IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF NORTH CAROLINA
CHARLOTTE DIVISION
No. 3:22-CV-0191-MOC-DCK

KANAUTICA ZAYRE-BROWN, Plaintiff, v. THE NORTH CAROLINA DEPARTMENT OF PUBLIC SAFETY, et al., Defendants.	)	**EXPERT REPORT OF FAN LI, PHD**

I. Introduction and Expert Background.

I was retained in this matter by Defendants to offer opinions and conclusions regarding the research related to the effectiveness of sex reassignment surgery (also known as gender affirming surgery) as a treatment for gender dysphoria. More specifically, I was asked to review studies referenced in the Standards of Care for the Health of Transgender and Gender Diverse People, Version 8 ("WPATH 8") which is published by the World Professional Association for Transgender Health ("WPATH") and the report of Plaintiff's expert, Randi Ettner, Ph.D., to assess their quality, review their research design, analyze their outcomes, and determine whether the studies support Plaintiff's and her expert's assertions. Additionally, I offer opinions and conclusions about this body of research in terms of study design and statistical methodology.

A. Summary Statement of Qualifications.

I am a statistician and a tenured Full Professor in the Department of Statistical Science, with a secondary appointment at the Department of Biostatistics and Bioinformatics at Duke University ("Duke"). I am a nationally leading expert on statistical methods for causal inference and comparative effectiveness research. I am the co-director of the Comparative Effectiveness

Methodology at the Duke Clinical Research Institute, which is the largest clinical research organization in the world. I have published extensively on statistical methods and applications in causal inference in leading professional peer-reviewed journals in medicine, statistics, epidemiology, including Journal of American Medical Association (JAMA), Proceedings of the National Academy of Sciences, American Journal of Epidemiology, and Journal of the American Statistical Association. I have been awarded multiple research grants from government agencies and research organizations, including the U.S. National Institutes of Health ("NIH"), National Science Foundation ("NSF"), Patient Centered Outcome Research Institute ("PCORI"), and industry sponsors. I was elected Fellow of the American Statistical Association in 2022 for my contribution to statistics research and service. I am serving as the editor for Social Science, Biostatistics and Policy for the premier statistics journal 'The Annals of Applied Statistics', and an associate editor for the premier statistics journal 'Journal of the American Statistical Association'. In addition, I have served as a reviewer for a large number of academic journals, including JAMA, Science, and grant proposal for funding agencies including the NIH and NSF. I have supervised five PhD dissertations, four postdoctoral fellows, and many masters and undergrad students at Duke. In addition to my research work, I have presented numerous seminars in national and international conferences and taught classes on causal inference at Duke. My research work in statistics focuses on causal inference, biostatistics, missing data, Bayesian analysis, and applications to health, policy, and social sciences.

I received my PhD in Biostatistics from Johns Hopkins University in 2006 and a Bachelor of Science in Mathematics from Peking University in China in 2001. I completed a postdoctoral fellowship in statistics in the Department of Health Care Policy at the Harvard Medical School in 2008. My curriculum vitae further sets forth my qualifications and is attached as Appendix A.

I have not previously done any research in the specific area of transgender health care. Nor have I been a member of WPATH, or otherwise been involved with the issue of transgender health care. This affords me the opportunity to engage in this project based solely on my education, training, and experience and without any preconceived notions of what the evidence may or may not show.

B. Information Considered in Forming Opinions.

In forming the conclusions and opinions set out in this report, I reviewed and considered a variety of materials and information, including the most recent WPATH Standards of Care, Version 8; the expert report of Randi C. Ettner, Ph.D., dated February 2, 2023, and her previous declarations; and the more than 80 studies relied upon by WPATH or Dr. Ettner, which are listed in Appendix B to this report.

In addition, I referenced the following studies when preparing this report:

1. Imbens GW and Rubin DB. 2015. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.
2. Hernan MA and Robins JM. 2023. Causal Inference: What If. CRC Press, Boca Raton.

C. Compensation, Other Expert Testimony, and List of Publications

For my work on this case, I am charging $350 per hour for all services that I render, except for time spent providing deposition or in-court testimony, for which I charge $475 per hour. I have not previously testified as an expert. My publications can be found on my CV, which is attached hereto as Appendix A.

II. Discussion

In this report, I summarize my opinions, provide background information regarding comparative effectiveness research and causal inference, and explain my methodology and approach. I have separately provided a detailed chart that analyzes the relevant studies that I reviewed in preparing this report. That chart is incorporated herein and is Appendix B to this report.

A. Summary of Opinions

Dr. Ettner and WPATH make several specific assertions regarding the effectiveness of gender affirming treatments, including surgical interventions, to treat gender dysphoria. Those assertions are identified in this report and in Appendix B. However, the evidence cited by Dr. Ettner and WPATH does not provide reasonable support for those assertions. Among the dozens of studies reviewed, there is not a single randomized controlled trial. Most of the studies cited in support of those assertions are of low quality in terms of study design and statistical methodology. Specifically, the majority of the studies are based on observational retrospective designs, which are prone to severe confounding bias. The few prospective studies that are cited point to mixed conclusions. Additional methodological shortcomings include small sample size, nonresponse bias, non-representative population (i.e., selection bias), self-reported outcomes. Moreover, most of the studies do not have before-after comparison of the same patients, and thus do not provide direct evidence on the effects of the treatment of interest. Also, the vast majority of these studies do not compare the results of sex reassignment surgery with alternative treatments, and thus do not provide evidence on the necessity or advantage of sex reassignment surgery over available alternative treatments. In summary, based on my education, training, and experience, I conclude to a reasonable degree of statistical certainty, that these studies fail to provide rigorous and

consistent statistical evidence on the benefits in quality of life and well-being of sex reassignment surgery (SRS).

B. Background Regarding Comparative Effectiveness Research and Causal Inference

In medicine, the type of research used to evaluate the effects and safety of an intervention is broadly referred to as comparative effectiveness research. The statistical methodology for quality of life belongs to the general statistical field of causal inference.

1. Association v. Causation

Statistics measures associations between variables based on data. A first lesson in elementary statistics is that association does not imply causation. The main barrier to interpreting the association between the treatment and the outcome as a causal effect is the presence of factors that are associated with both the treatment and the outcome. These factors are commonly referred to as confounders or confounding variables or confounding factors. For example, patients with worse health conditions may be more likely to obtain a beneficial medical treatment. So directly comparing the outcomes of the treated and control patients, without adjusting for the difference in their baseline health conditions, would bias the causal comparisons and mistakenly conclude that the treatment is harmful. This type of bias is called confounding bias.

In the context of medical care for the transgender population, a hypothetical example is that transgender patients who received sex reassignment surgery (SRS) may be systematically different from the transgender patients who did not receive SRS, perhaps in terms of confounders such as physical and mental conditions or social economics status. So, the difference in the outcome (e.g., quality of life) between transgender patients who received SRS and those who did not receive SRS may be due to the difference in the confounders between these two groups of transgender patients.

When confounders are observed and measured, analysts can use statistical methods for causal inference, e.g., multivariable regression, propensity scores, matching (Imbens and Rubin, 2015; Hernan and Robins, 2020), to control for the bias due to the difference in the confounders between treatment and control groups. When confounders are unobserved or unmeasured, statistical analysis on the causal effect of a treatment based on the observable data is subject to confounding bias.

2. Types of Biases in Causal Interference

There are several types of bias that must be accounted for in assessing research regarding the efficacy of medical interventions. The primary types of bias are set forth and explained below.

a. Confounding Bias

As explained above, confounding bias occurs when the treatment and control groups of individuals differ in observed or unobserved factors that can also affect the outcome. Randomized controlled trials eliminates all confounding bias, and therefore the association between treatments and outcomes can be interpreted as causal effects. In contrast, observational studies, regardless of the statistical analytical methods used, cannot rule out confounding bias. Therefore, in order to interpret the association between treatments and outcomes as causal effects in observational studies, one must assume that there is no unmeasured confounding factor. Such an assumption is untestable and is almost always untenable.

b. Baseline Outcome

When a confounder is strongly predictive of the outcome(s), failing to adjust for it could induce particularly large confounding bias. In comparative effectiveness research, one of the most important confounders is the baseline outcome, i.e., the outcome measured before the treatment. For example, in studying the effect of SRS on quality of life, an important confounder is the pre-

operative measurement of quality of life. A comparison of the outcome of the same subject before and after the treatment would eliminate the confounding bias due to the difference in the baseline outcomes. In this example, this could be done through a paired t-test or a multivariable regression with the before-after difference of the outcome as the dependent variable.

c. Selection Bias

Selection bias occurs when participants of a study are selected in a way that does not make them representative—in observed or unobserved ways—of the population to which the study purports to apply. A special case of selection bias is attrition bias, which occurs when patients are lost to follow up in a longitudinal study as these patients may be systematically different from the patients who stayed in the study.

d. Nonresponse Bias

Nonresponse bias occurs when the response rate of a survey is less than 100 percent and those who do not respond to a survey may have answered differently than those who do respond. This can create biased results that do not accurately reflect the population of interest. The lower the response rate is, the more likely severe nonresponse bias occurs. Nonresponse bias is a special type of selection bias. See A Systematic Review of Nonresponse Bias.

e. Recall Bias

Recall bias occurs when participants in a research study do not accurately remember a past event or experience or leave out details when reporting about them. It can occur in studies that ask participants to provide information from memory, such as in case-control studies or retrospective cohort studies. See recall-bias.

3. Study Designs in Comparative Effectiveness Research

Understanding the type of study design is essential to assessing the strength of any conclusion regarding the effectiveness of a medical intervention.

a. Randomized Controlled Trials Versus Observational Studies

In medicine, the consensus gold standard for evaluating the efficacy, effectiveness, and safety of an intervention or treatment is the randomized controlled trials. For example, the U.S. Food and Drug Administration (FDA) requires evidence of efficacy and safety of a new product based on randomized controlled trials in the approval of the vast majority of new drugs and medical devices. When randomized controlled trials are not available, researchers resort to observational studies for comparative effective research.

In a randomized controlled trial, a sample of patients are randomly assigned to either a treatment or a control group, and after the treatment or control condition is applied, the outcomes of interest—also known as dependent variables—are measured and then compared. In addition, information on patients' baseline characteristics, e.g., age, sex, health conditions, is routinely collected before the treatment. Such factors are often referred to as prognostic factors, or pre-treatment variables, or baseline covariates. Randomized controlled trials are the gold standard for causal inference because it eliminates all confounding bias due to both measured and unmeasured confounders.

In observational studies, the process of determining which patients receive which treatment condition is unknown and uncontrolled by the researchers. The treated and control subjects may differ in their baseline characteristics. If these baseline characteristics are confounders, then failing to statistically adjust for them would induce bias in evaluating the causal effect of the treatment. In observational studies, there is no guarantee that the researchers observed all confounders. So,

in order to interpret the association between a specific treatment and outcomes as causal effect, one must make the untestable assumption that there is no unmeasured confounding factor. This assumption is likely violated to a certain degree of confounding factor(s) in most observational studies (Imbens and Rubin, 2015; Hernan and Robins, 2020). This is why observational studies are inferior to randomized controlled trials in causal inference.

b. Prospective Studies Versus Retrospective Studies (Types of Observational Studies)

In a prospective or prospective cohort study, the treatment of interest (e.g., a sexual reassignment surgery) has not occurred at the time the study is initiated, and the study subjects are followed with the information being collected at one or multiple time points during the study period. In contrast, in a retrospective study, the treatment of interest has already occurred at the time the study is initiated, and the study subjects are identified from existing data, e.g., medical records, and information is collected about their past.

Prospective designs are generally considered superior to retrospective designs, because prospective studies collect information as the subjects' characteristics or circumstances change whereas retrospective studies collect static information. Moreover, retrospective studies are more likely subject to recall bias. Note that randomized controlled trials are prospective studies, but not all prospective studies are randomized controlled trials.

c. Longitudinal Studies Versus Cross-Sectional Studies

A longitudinal study measures the same outcome of each subject multiple times over a period of time. A cross-sectional study measures the outcome of each subject only one time. Longitudinal studies are superior to cross-sectional studies because they generally provide more information, particularly on the trend or progression of an outcome. Both randomized controlled trials and observational studies can be either cross-sectional or longitudinal.

d. Before-After Study

An important type of prospective study is the before-after study, where the outcomes of the same patients are measured both before and after the treatment. A before-after comparison of the outcome eliminates the confounding bias due to the difference in the baseline outcomes, which are often the most important confounding factors.

4. Quality of the Designs

For causal inference, the best design is a randomized controlled trial, which eliminates all confounding bias. The next preferred design is a prospective observational study with outcomes measured both before and after the intervention, which, in combination with proper statistical methods, can eliminate confounding bias due to measured confounders, but cannot eliminate confounding bias due to unmeasured confounders. The design of the lowest quality is a retrospective observational study, which is prone to confounding bias from both measured and unmeasured confounders, particularly when the baseline outcomes are not recorded.

****

C. Assessment of Research Referenced by WPATH and Dr. Ettner

Throughout this report and in Appendix B, I use various abbreviations. SRS refers to sex reassignment surgery (SRS). The literature variably refers to surgical interventions as SRS, gender-affirming surgery, or gender-reassignment surgery. I use SRS for consistency. I use QoL for quality of life and GAT for gender-affirming treatments. GAT includes various interventions, which can include SRS or hormonal treatments.

1. Analysis of Plaintiff's Contentions

I have carefully reviewed and assessed Dr. Ettner and WPATH's conclusions and assertions on the evidence regarding SRS and find these assertions to not be well founded for the reasons discussed below. My investigation largely focuses on the specific assertions that relate to the efficacy of SRS and that purport to be supported by research. Appendix B is a table that collects and summarizes my analysis of these assertions and the studies referenced therein. The italicized text which appears after each numbered assertion below is a direct quote from the source material that forms the basis of the assertion analyzed.

2. Analysis of Key Assertions

In this section, I discuss select assertions made by WPATH and Dr. Ettner and analyze the studies cited to support those assertions. See Appendix B for a complete listing of the assertions and studies analyzed.

WPATH Assertion 1. *There is strong evidence demonstrating the benefits in quality of life and well-being of gender-affirming treatments, including endocrine and surgical procedures, properly indicated and performed as outlined by the Standards of Care (Version 8), in TGD people in need of these treatments.* (WPATH 8 at S18)

My examination of this assertion focused on the statement "there is strong evidence demonstrating the benefits in quality of life and well-being of gender-affirming treatments." This assertion cites 21 references. Given the number of studies cited in support of this assertion, I summarize my review of these studies by the design study and paper types. Note that not all these studies focus on SRS; many studies investigate non-surgery treatments such as hormone therapy (e.g., Wierckx, K., Van Caenegem, et al. 2014; T'Sjoen et al., 2019; Baker et al. 2021). Therefore, below I refer to the treatments in the studies cited in this assertion generically as gender-affirming treatments (GAT).

Among all the 21 studies reviewed relative to this assertion, there is not a single randomized control trial. In fact, in all the literature reviews cited here, only one randomized control trial (Pelusi et al., 2014) was mentioned, which involves 45 patients and focuses on hormonal therapy, but this paper is not cited in WPATH Assertion 1. So, all the studies cited in this assertion are subject to confounding bias.

Summary of the Prospective Studies Cited in Support of WPATH Assertion 1. There are five prospective studies (Wierckx, K., Van Caenegem, et al. 2014; Cardoso da Silva et al., 2016; Lindqvist et al, 2017; van de Grift, Elaut et al., 2018; Aires et al., 2020). Among these studies, Lindqvist et al. (2017) has the best study design, which is a prospective cohort study with before-surgery outcome and after-surgery outcome measured at multiple time point. Lindqvist et al. (2017) found mixed results on the effects of SRS on QoL. Specifically, Lindqvist et al. (2017) found that compared to before the treatment, QoL is better one year after operation but worse 3 and 5 years after operation.

Another prospective study with before-after outcomes (Cardoso da Silva et al., 2016) also found mixed results. Specifically, Cardoso da Silva et al., (2016) found that psychological health

and social relationships were significantly improved after SRS, but physical health and level of independence were significantly worse after SRS. Another prospective study (Aires et al., 2020) focuses on a special type of surgery of chondrolaryngoplasty or tracheal shaving and its effects on voice and visual aesthetic satisfaction. That study provides no information about the effects of the surgery on general QoL and well-being of the patients. Wierckx, K., Van Caenegem, et al. (2014) is a prospective study focusing on hormonal therapy, specifically it focuses on the safety and side effects of the intervention, not QoL. Despite being a prospective study, van de Grift, Elaut et al., 2018 does not provide before-after comparison of the same patients. In short, the prospective studies, at best, provide mixed results on the effects of SRS.

Summary of Retrospective Studies Cited in Support of WPATH Assertion 1. There are nine retrospective studies (Ainsworth & Spiegel, 2010; Buncamper et al., 2016; Yang, Zhao et al., 2016; Owen-Smith et al., 2018; Özkan et al., 2018; Al-Tamimi et al., 2019; Balakrishnan et al., 2020; Almazan & Keuroghlian, 2021; Mullins et al. 2021). None of these studies provide direct before-after comparison in QoL or well-being of the same patients. The comparisons are usually between people who received gender-affirming treatments and those who did not. So, this study does not provide direct evidence on the effect of GAT on individual patients, and is subject to severe confounding bias. In addition, four studies (Al-Tamimi et al., 2019; Buncamper et al., 2016; Özkan et al., 2018; Balakrishnan et al., 2020) are descriptive analyses (e.g., describing surgical techniques) and did not provide any type of comparison. Mullins et al. (2021) focuses on thrombosis and thrombosis risk factors among adolescent and young adult transgender population and does not discuss QoL.

Summary of Literature Reviews Cited in Support of WPATH Assertion 1. There are seven literature reviews (Poteat et al., 2016; White Hughto and Reisner, 2016; Nobili et al., 2018;

T'Sjoen et al., 2019; Eftekhar Ardebili et al. 2020; Baker et al. 2021; Javier et al., 2022). One study (Poteat et al. 2016) reviews the HIV epidemic among transgender population and does not directly discuss QoL and well-being related to GAT. Three studies (White Hughto & Reisner 2016; T'Sjoen et al., 2019; Baker et al. 2021) focus on the effects of hormonal therapy. One study (Nobili et al., 2018) reviews a combination of surgery and hormonal therapies. One study (Javier et al., 2022) found mixed results on the effects of GAT. Two studies (Nobili et al., 2018; Eftekhar Ardebili et al. 2020) provide meta-analysis of multiple studies, but do not provide information on the effects of GAT on QoL because of the lack of before-treatment outcomes.

Most of the review papers acknowledge the methodological shortcomings of the studies in this field. For example, Nobili et al. (2018) acknowledges "the majority of the studies were cross-sectional, lacked controls, and displayed moderate risk of bias…Better quality studies that include clearly defined transgender populations, divided by stage of gender affirming treatment and with appropriate matched control groups are needed to draw firmer conclusions." Baker et al. (2021) acknowledges "this conclusion is limited by high risk of bias in study designs, small sample sizes, and confounding with other interventions." T'Sjoen et al. (2019) acknowledges "current available research is based mostly on cross-sectional studies…long-term follow-up studies and studies involving large groups of people are needed to evaluate whether these improvements remain." Javier et al. (2022) acknowledges their review is based on "seventy-nine low quality studies." White Hughto & Reisner, Poteat et al. (2016) comments "prospective controlled trials are needed to investigate the effects of hormone therapy on the mental health of transgender people."

Overall Assessment of WPATH Assertion 1. Most of the studies cited in support of this assertion are of low quality in terms of study design and statistical methodology. There is not a single randomized controlled trial. All studies are observational studies. Among these, the few

prospective studies (which are of higher quality of retrospective studies) point to mixed conclusions regarding the effects of GAT on QoL and well-being of the transgender patients. Most of the studies are subject to confounding bias due to the potential difference between the transgender population who received GAT and those who did not. Most of the studies do not have before-after comparison of the same patients and thus do not provide direct evidence on the effects of GAT for individuals. In addition, none of the studies compare SRS with alternative treatments such as hormone therapy or compare between different types of SRS, and thus do not provide evidence for the comparative advantage of SRS over alternative treatments or one specific type of SRS over another. Moreover, most of the studies suffer from additional methodological shortcomings such as small sample size, nonresponse bias, non-representative population. In addition, often the outcomes are not measured using standardized and validated instrument. In summary, I conclude that contrary to the statement in the assertion, these studies fail to provide rigorous and consistent statistical evidence on the benefits in quality of life and well-being of gender-affirming treatments.

WPATH Assertion 2. *Gender-affirming interventions are based on decades of clinical experience and research; therefore, they are not considered experimental, cosmetic, or for the mere convenience of a patient. They are safe and effective at reducing gender incongruence and gender dysphoria.* (WPATH 8 at S18).

My assessment of this assertion focuses on the statement of "they are … effective at reducing gender incongruence and gender dysphoria." This is a statement on the comparative effectiveness of GAT because the term "reducing" indicates comparisons. This assertion cited 25 references, 15 of which were cited also in WPATH Assertion 1.

First of all, not a single study directly measures gender incongruence and gender dysphoria as the main outcome variable but instead used derivative measures, for example, satisfaction with surgery or quality of life in general. This shortcoming was also pointed out in Gijs and Brewaeys (2007), a reference cited in support of Ettner Assertion 2, which is discussed below. As a result, none of the studies provide direct evidence on the effects of GAT on gender incongruence and gender dysphoria.

Second, the study designs of these studies are generally of low quality. There is not a single randomized controlled trial, and all studies are observational studies. Among these observational studies, there are only four prospective studies with before-after comparison of the same patients (Wierckx, K., Van Caenegem, et al. 2014; Lindqvist et al. 2017; Aires et al., 2020; Aldridge et al., 2020), two of which (Lindqvist et al. 2017; Aldridge et al., 2020) reported mixed results in whether GAT improves various measurements of mental health, and two (Wierckx, K., Van Caenegem, et al. 2014; Aires et al. 2020) do not provide data on outcomes related to mental health or QoL. Specific to this assertion, it focuses on hormone treatments and provides no information on the effects of SRS. One paper (White Hughto, Reisner, Poteat et al., 2016) is a literature review of three prospective studies of the effects of hormone therapy on psychological functioning and QoL, and it calls for "prospective controlled trials." None of the rest 20 studies provide before-after comparisons of any outcome. In fact, seven of these papers focus on describing the surgical techniques of various SRS (Buncamper et al., 2016; Lo Russo et al. 2017; Wolter et al. 2015, 2018; Claes et al., 2018; Esmonde et al., 2019; Balakrishnan et al., 2020) and do not discuss mental health or QoL.

In summary, due to low quality study design, all the studies cited in support of WPATH Assertion 2 are subject to confounding bias, and most are subject to selection bias, nonresponse

bias and recall bias. Thus, it is my opinion that the cited references fail to provide rigorous statistical evidence to support the assertion that "gender-affirming interventions are … effective at reducing gender incongruence and gender dysphoria."

WPATH Assertion 6. *Controlled studies show clinically significant health and mental health disparities for justice-involved transgender people compared to matched groups of transgender people who have not been incarcerated or jailed.* (Brown and Jones, 2015). WPATH S104

Brown and Jones (2015) compares health disparities between transgender and non-transgender veterans, and found transgender veterans have an increased likelihood of justice involvement. This comparison is entirely different from the comparison stated in the assertion—regarding health disparities of justice-involved transgender people vs. non-justice-involved transgender people (i.e., comparison between different transgender populations). Therefore, Brown and Jones (2015) is not relevant to and does not support this assertion.

WPATH Assertion 10. *Gender-affirming vaginoplasty is one of the most frequently reported gender-affirming surgical interventions […] Although different assessment measurements were used, the results from all studies consistently reported both a high level of patient satisfaction (78–100%) as well as satisfaction with sexual function (75–100%). Although different assessment measurements were used, the results from all studies consistently reported both a high level of patient satisfaction (78–100%) as well as satisfaction with sexual function (75 100%). This was especially evident when using more recent surgical techniques. Gender affirming vaginoplasty was also associated with a low rate of complications and a low incidence of regret (0–8%)*. (WPATH 8 at S128-S129)

This assertion cites 26 references, 3 of which were cited also in WPATH Assertion 1. Among all the studies cited, there is not a single randomized controlled trial on the safety and effects of gender-affirming vaginoplasty (GAV). All studies are observational studies. Among these, there are only two prospective studies with before-after surgery comparisons: Papadopulos, Zavlin et al. (2017) evaluate a combined SRS technique instead of GAV alone and focused on short-term (6-month after surgery) outcomes; Cardoso da Silva et al. (2016) found mixed results in that psychological health and social relationships were significantly improved after SRS, but physical health and level of independence were significantly worse after SRS.

The rest of the 26 references do not provide before-after comparison or comparison between transgender patients who underwent GAV and who did not. Among the retrospective studies, Simonsen et al. (2016) is the one of the highest quality because it uses a national (Denmark) registry to investigate the entire Danish transgender population in 32 years. The results are inconclusive, as the authors concluded "generally SRS may reduce psychological morbidity for some individuals while increasing it for others." Buncamper et al. (2015) found high rate of sexual dysfunctionality following SRS. Most studies use self-reported outcomes instead of standardized instruments. Lastly, most studies do not discuss QoL outcomes.

In summary, the cited references suffer from several methodological shortcomings, including lack of randomized controlled trials and even prospective before-after studies (and thus subject to confounding bias), self-reported outcomes (e.g., satisfaction), nonresponse bias. None of the studies compare GAV with alternative treatments. This body of literature supports the high self-reported satisfaction rate among the patients who underwent GAV but does not provide any evidence for the necessity or advantage of GAV comparing to alternative treatments.

WPATH Assertion 11. *Gender-affirming surgical procedures have been shown to relieve symptoms of gender dysphoria and improve mental health.* (WPATH 8 at S173)

This assertion cites two references (van de Grift, Elaut et al., 2017; Owen-Smith et al. 2018), both of which were cited in previous Assertions. Neither study provides before-after SRS comparison of the same patients, and thus is subject to severe confounding bias, nor do they provide any direct evidence for the assertion that these procedures "relieve symptoms of gender dysphoria and improve mental health."

Ettner Assertion 1. *Decades of careful and methodologically sound scientific research have demonstrated that gender-affirming surgeries are safe and effective treatments for severe gender dysphoria and, indeed, for many people suffering from gender dysphoria, the only effective treatment.* (Ettner Report ¶ 50)

This assertion cited three references. None of the references provides evidence for the assertion that "for many people suffering from gender dysphoria, the only effective treatment." In fact, Pfafflin and Junge (1998) states "Sex reassignment…, is, however, *not the only powerful change agent in sex reassignment*." As elaborated in my assessment of WPATH Assertion 1, the statistical methodology in the field of comparative effectiveness of SRS is not up to the long-established standard in comparative effectiveness research in medicine. There has not been a single randomized control trial on SRS. All cited studies are observational studies. Among the observational studies, the vast majority of the studies are cross-sectional retrospective design instead of the prospective before-after design. Many of the methodological shortcomings have been repeatedly pointed out in numerous literature reviews in this field (e.g., several cited in WPATH Assertion 1). Therefore, I, as an expert who specializes in statistical methodology for

comparative effectiveness research, strongly disagree with Dr. Ettner's characterization of this research as "methodologically sound scientific research."

Ettner Assertion 2. *In 2007, Gijs and Brewayes analyzed 18 studies published between 1990 and 2007, encompassing 807 patients. The researchers concluded: "Summarizing the results from the 18 outcome studies of the last two decades, the conclusion that [gender-affirming surgery] is the most appropriate treatment to alleviate the suffering of extremely gender dysphoric individuals still stands: Ninety-six percent of the persons who underwent [surgery] were satisfied and regret was rare."* (Ettner Report ¶ 55)

Immediately after the quote contained in the above assertion (p. 215), the authors (Gijs and Brewayes) acknowledged that "[h]owever, even today this conclusion is based on methodologically less than perfectly designed studies." Specifically, the paper wrote "[n]ot one of the reviewed outcome studies was a controlled one...In many studies, sound psychometric instruments were not used. Especially disturbing is that many researchers did not directly measure gender dysphoria as the main outcome variable but instead used derivative measures, for example, satisfaction with surgery, sexual and interpersonal relationships, occupational and global functioning, or quality of life in general." The authors also acknowledge a few other methodological shortcomings, like attrition or selection bias of the patient sample, which echo my critiques with respect to WPATH Assertion 1.

Ettner Assertion 4. *Studies conducted in countries throughout the world likewise conclude that gender-affirming surgery is an extremely effective treatment for gender dysphoria.* (Ettner Report ¶ 57)

This statement overlaps with WPATH Assertions 1 and 2. As has already been discussed in the assessments of those Assertions, as well as to Ettner Assertion 1, it is my opinion that the

statistical analysis of the body of literature on the effectiveness of SRS is not methodologically sound and has failed to provide rigorous and consistent scientific evidence for the benefits of gender-affirming treatments on reducing gender dysphoria or improving and well-being of the patients.

Ettner Assertions 5-9. Assertions 5-9 are assessed together because they all concern the effectiveness of SRS. Collectively the assertions read: *Studies have shown that by alleviating the suffering and dysfunction caused by severe gender dysphoria, gender-affirming surgery improves virtually every facet of a patient's life. This includes satisfaction with interpersonal relationships and improved social functioning, […] improvement in self-image and satisfaction with body and physical appearance and greater acceptance and integration into the family. […] Studies have also shown that gender-affirming surgery improves patients' abilities to initiate and maintain intimate relationships.* (Ettner Report ¶¶ 58-59)

First, among the 24 cited references (half were cited in previous assertions), there is not a single randomized controlled trial. All studies are observational studies. Among these, there are only two prospective studies with before-after SRS comparisons (Smith et al. 2005; Johansson et al. 2009), and the rest are retrospective studies without not before-after comparison of the same patients.

Second, the conclusions from the cited references showed great variability in the outcomes and is not as consistent and clear cut as indicated in the Assertion. For example, Klein and Gorzalka (2009) is a literature review and it states, "From this review it is clear that there is great variability in the sexual functioning of post-operative transsexuals, and that no clear outcome can be predicted with respect to whether surgery will have a positive or negative impact on sexual function." Weyers et al. (2009) found "they suffer from specific difficulties, especially concerning arousal,

lubrication, and pain." Lobato et al. (2006) found that family relationship only improved in 1/4 of the patients. In addition, some of the papers do not report on outcomes concerning the specific assertion, e.g., Lawrence 2003 for Supplemental Assertion 6; Jarolím et al. 2009, Rehman et al. 1999 for the assertion regarding improved intimate relationships.

Third, two studies (Johansson et al. 2009, 2010) found a significant discrepancy between clinicians and patients' assessment of the same outcome, with patients reporting a higher favorable rate (e.g., 33% in Johansson et al. 2009) than the clinicians. This raises concerns as to the accuracy and reliability of the patient self-reported outcomes.

In summary, similar to my assessment of comparable assertions (namely, WPATH Assertions 1, and 2, Ettner Assertions 1, 2, and 4), it is my opinion that this body of literature on the effectiveness of SRS cited in Ettner Assertion 5-9 is not methodologically sound and has failed to provide rigorous and consistent scientific evidence for the benefits of gender-affirming treatments on reducing gender dysphoria or improving and well-being of the patients.

Ettner Assertion 10. *Research shows that the risk of suicide can be significantly diminished with prompt and effective treatment*. (Ettner Report ¶ 77)

I carefully reviewed the part of the paper that is relevant to the assertion. Specifically, the paper states (page 12 of 15), "The process of medically transitioning overall was more complex…We did not observe an increased risk in this sub-group among those who completed a medical transition (RR = 0.51; 0.07, 3.74)." Exact numbers are reported on the second page of Table 4 (page 11 of 15, top row); it shows that among the subjects who completed medical transition (100 subjects), the relative risk of suicidal attempt is 0.51 with 95% confidence interval (0.07, 3.74), which is ***not*** statistically significant. So, the claim of "significantly diminished" is not supported by the statistical analysis in this reference.

Ettner Assertion 11. *Gender dysphoric individuals have a profound discomfort or disgust of their genitalia. Without effective treatment as outlined above, this often leads to attempts at surgical self-treatment (SST), which can result in lasting physical trauma or death*. (Ettner Report ¶ 78)

This assertion cites Brown and McDuffie (2009), which does not provide any information on the frequency of this behavior. Moreover, on the bottom row of Page 187, the authors write "this (surgical self-treatment) rarely occurs in the community absent psychosis[.]" This is directly at odds with the sweeping assertion of that without surgery dysphoric individuals "often" engage in such behavior. During my review, I found Brown (2014), which reported that "five percent of [TG] inmates reported that they had attempted (2%) or completed (3%) autocastration while incarcerated." This percentage supports the occurrence is relatively rare rather than often. Moreover, Brown (2014) uses a highly selected sample of the letters written by transgender inmates—the representativeness of that sample of the general incarcerated transgender population is unclear.

Ettner Assertion 12. *[A] systematic meta-analysis on publications performed by German researchers included 1,100 post-surgery participants. Seven different measures of quality of life were employed. The researchers concluded that gender-affirming surgery positively affects well being, sexuality, and quality of life in general.* (Ettner Report ¶ 116)

The referenced study is Weinforth et al. (2019), which is a literature review of 13 studies. These studies adopt mixed designs (some prospective and some retrospective) and different measurements for outcome, and most studies focus on short term outcomes. Weinforth et al. (2019) only provided qualitative summaries, not any formal statistical meta-analysis. Only two studies are prospective (Cardoso da Silva et al.; Lindqvist et al.), which have been reviewed in my

assessment of WPATH Assertion 1, and both showed mixed to negative results. The other 9 studies are all of low quality; the methodological problems include retrospective cross-sectional studies, small sample sizes, no causal inference methods – no formal confounding adjustment other than regression. In fact, the authors acknowledged the methodological shortcomings and noted that "prospective studies with standardized methods of assessing quality of life and with longer follow-up times would be desirable." Similar to my assessment of the previous assertions, I determine that this reference fails to provide rigorous and consistent statistical evidence supporting the benefits of SRS.

3. Description of Appendix B – Table of Reviewed Studies

In preparing this report, I reviewed various portions of WPATH 8 and Dr. Ettner's expert report which contained assertions regarding the body of research on the effectiveness of various treatments, including surgery, in treating gender dysphoria and/or improving well-being. I reviewed the studies that were cited to support 11 assertions made in WPATH 8 and 12 assertions made by Dr. Ettner in her report. Those assertions are labeled in Appendix B as WPATH Assertion 1, 2, and so on, and Ettner Assertion 1, 2 and so on. Immediately beneath the assertion label is the text of that assertion, including the cited studies, along with the page reference. Below the assertion text, one row is dedicated to each study and for each the following summary information placed in corresponding columns:

- Study-design – category of study
- Methodology – brief description of how the study was conducted
- Aim/Objective – brief statement of the question the researchers sought to examine
- Conclusion – summary of the findings
- Strengths – self-explanatory

- Limitations – self-explanatory
- Notes – key takeaways, if any.

III. Conclusion

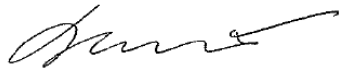
The studies that Dr. Ettner and WPATH rely upon in making their assertions regarding the effectiveness of SRS and other treatments do not constitute a rigorous and consistent body of statistical evidence. These studies are of low quality in terms of study design and statistical methodology, due to issues like confounding bias, small sample sizes, nonresponse bias, non-representative population, more. Additionally, most of the studies do not have before-after comparison of the same patients and thus do not provide direct evidence on the effects of the treatment of interest. Moreover, the vast majority of these studies do not compare the results sex reassignment surgery with alternative treatments, and thus do not provide evidence on the necessity or advantage of sex reassignment surgery over available alternative treatments. Thus, it is my opinion, to a reasonable degree of statistical certainty that the studies cited by Ettner and/or WPATH and reviewed in this report simply do not provide reasonable support for the assertions made by Dr. Ettner and WPATH relative to the benefits in quality of life and well-being of gender-affirming treatments.

THIS SPACE IS INTENTIONALLY LEFT BLANK

SIGNATURE PAGE TO FOLLOW

I, Fan Li, pursuant to 28 U.S.C. § 1746, declare that the foregoing is true and correct.

This the 17th, day of June, 2023.



Fan Li, PhD

May 24, 2023

CURRICULUM VITAE

Fan Li

Department of Statistical Science
Duke University Box 90251
Durham, NC 27708
Email: fl35@duke.edu
Webpage: https://www2.stat.duke.edu/~fl35/

EDUCATION

2006	Ph.D., Biostatistics, Johns Hopkins University
2001	B.Sc., Mathematics, Peking University, China

POSTDOCTORAL TRAINING

2006-2008	Postdoctoral Fellow in Statistics Department of Health Care Policy, Harvard Medical School

PRIMARY ACADEMIC APPOINTMENT

(All in Department of Statistical Science, Duke University)

2021-present	Professor
2015-2021	Associate Professor
2008-2015	Assistant Professor

SECONDARY ACADEMIC APPOINTMENT

2021-present	Professor Department of Biostatistics and Bioinformatics, Duke University
2017-2021	Associate Professor Department of Biostatistics and Bioinformatics, Duke University
2018-present	Co-director Program for Comparative Effectiveness Methodology, Duke Clinical Research Institute
2017-present	Affiliated Faculty Duke Clinical Research Institute

HONORS AND AWARDS

2022 Fellow, American Statistical Association

PUBLICATIONS

Peer-reviewed Articles

(* student or postdoc supervised by FL)

1. **Li F**, and Frangakis CE (2005). Designs for partially controlled studies: Messages from a review. *Statistical Methods in Medical Research*, 14, 417-431.

2. **Li F**, and Frangakis CE (2006). Polydesigns and causal inference. *Biometrics*, 62(2), 343-351.

3. Baccini M, Cook S, Frangakis CE, **Li F**, Mealli F, Rubin DB, and Zell EZ. (2010). Multiple imputation in the Anthrax Vaccine Research Program. *Chance*, 23(2), 16-23.

4. **Li F**, Green JG, Zaslavsky AM, and Kessler R. (2010). Estimating prevalence of serious emotional disturbance in schools using a brief screening scale. *International Journal of Methods in Psychiatric Research*, 19 (Supplement 1), 88-98.

5. **Li F**, and Zhang NR. (2010). Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *Journal of the American Statistical Association*, 105(491), 1202-1214.

6. **Li F**, and Zaslavsky AM. (2010). Using a short screening scale for small-area estimation of mental illness prevalence for Schools. *Journal of the American Statistical Association*, 105(492), 1323-1332.

7. Schwartz SL*, **Li F**, and Mealli F. (2011). A Bayesian semiparametric approach to intermediate variables in causal inference. *Journal of the American Statistical Association*, 106(496), 1331-1344.

8. Go VF, Frangakis CE, Nam LV, Sripaipan T, Bergenstrom A, **Li F**, Latkin, C, Celentano, DD, and Quan, VM. (2011). Characteristics of high risk HIV-positive IDUs in Vietnam: implications for future interventions. *Substance Use and Misuse*, 46(4), 381-389.

9. Schwartz SL*, **Li F**, and Reiter JP. (2012). Sensitivity analysis for unmeasured confounding in principal stratification. *Statistics in Medicine*, 31(10), 949-962.

10. Zhang T, **Li F**, Beckes L, Brown C, and Coan JA. (2012). Nonparametric inference of hemodynamic response using multi-subject fMRI data. *NeuroImage*, 63, 1754-1765.

11. Zhang T, **Li F**, Beckes L, and Coan JA. (2013). A semi-parametric model of the hemodynamic response for multi-subject fMRI data. *NeuroImage*, 75, 136-145. (featured in NSF highlight 24408 “Reach out and touch someone”)

12. **Li F**, Zaslavsky AM, and Landrum MB. (2013). Propensity score weighting with multilevel data. *Statistics in Medicine*, 32(19), 3373-3387.

13. Mattei A, **Li F**, and Mealli F. (2013). Exploiting multiple outcomes in Bayesian principal stratification analysis with application to the evaluation of a job training program. *Annals of Applied Statistics*, 7(4), 2336-2360.

14. Liu F, Chakraborty S, **Li F**, Liu Y, and Lozano AC. (2014). Bayesian regularization via Graph Laplacian. *Bayesian Analysis*, 9(2), 449-474.

15. Zhang T, **Li F**, Gonzalez M, Maresh E, and Coan JA. (2014). A semi-parametric nonlinear model for event-related fMRI. *NeuroImage*, 97, 178-187.

16. **Li F**, Baccini, M, Mealli, F, Zell, EZ, Frangakis, CE, and Rubin, DB. (2014). Multiple imputation by ordered monotone blocks with application to the Anthrax Vaccine Research Program. *Journal of Computational and Graphical Statistics*. 23(3), 877-892.

17. **Li F**, and Mealli, F. (2014). A conversation with Donald B. Rubin. *Statistical Science*. 29(3), 439-457.

18. Mercatanti, A, and **Li F**. (2014). Do debit cards increase household spending? Evidence from a semiparametric causal analysis of a survey. *Annals of Applied Statistics*. 8(4), 2405-2508.

19. Schliep, EM, Dong, Q, Gelfand, AE, and **Li F**. (2014). Modeling individual tree growth fusing diameter tape and increment core data. *Environmetrics*. 25(8), 610-620.

20. Mercatanti, A, **Li F**, and Mealli, F. (2014). Improving inference of Gaussian mixtures using auxiliary variables. *Statistical Analysis and Data Mining*. 8(1), 34-48.

21. Zhang, T, Wu, J, **Li F**, Boatman-Reich, D, and Caffo, B. (2015). A Directional dynamic model for effective brain connectivity using electrocorticographic (ECoG) time series. *Journal of the American Statistical Association*. 110(509), 93-106.

22. **Li F**, Zhang T, Wang Q, Gonzalez M, Maresh E, and Coan JA. (2015). Spatial Bayesian variable selection and grouping in high-dimensional scalar-on-image regressions. *Annals of Applied Statistics*. 9(2), 687-713.

23. **Li F**, Mattei A, and Mealli F. (2015). Evaluating the effect of university grants on student dropout: Evidence from a regression discontinuity design using Principal Stratification. *Annals of Applied Statistics*. 9(4), 1906-1931.

24. Mercatanti A, and **Li F**. (2017). Do debit cards decrease cash demands?: Causal inference and sensitivity analysis using Principal Stratification. *Journal of Royal Statistical Society - Series C (Applied Statistics)*. 66(4), 759-776. (selected by the Royal Statistical Society (RSS) editors to present at the 2018 RSS Conference)

25. Akande O*, **Li F**, and Reiter JP. (2017). An empirical comparison of multiple imputation methods for categorical data. *American Statistician*. 71(2), 162-170.

26. Wang F, Wang J, Gelfand AE, and **Li F**. (2017). Accommodating the ecological fallacy in disease mapping in the absence of individual exposures. *Statistics in Medicine*. 36, 4930-4942.

27. Brennan JM, Thomas LE, et al., **Li F**, E Petersen. (2017). Transcatheter Versus Surgical Aortic Valve Replacement: Propensity-Matched Comparison. *Journal of American College of Cardiology*. 70, 439-450.

28. **Li F**, Morgan KL, and Zaslavsky AM. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*. 113(521), 390-400.

29. Ding P, and **Li F**.(2018). Causal inference: a missing data perspective. *Statistical Science*. 33(2), 214-237.

30. Kaufman BG, Klemish D, Kassner C, Reiter JP, **Li F**, Harker M, O'Brien EC, Taylor D, Bhavsar N. Predicting Length of Hospice Stay: An Application of Quantile Regression. (2018). *Journal of Palliative Medicine*. 21 (8), 1131-1136.

31. Arnold SV, Cohen DJ, Dai D, Jones PG, **Li F**, Thomas L, Baron SJ, Frankel NZ, Strong S, Matsouaka RA, Edwards FH, Brennan JM. (2018). Predicting Quality of Life at 1 Year after Transcatheter Aortic Valve Replacement in a Real-World Population. *Circulation: Cardiovascular Quality and Outcomes*. 11(10), e004693.

32. Wang F, Wang J, Gelfand AE, and **Li F**. (2019). Disease mapping with generative models. *American Statistician*. 73(3), 212-223.

33. Li F*, Thomas LE, and **Li F**. (2019). Addressing extreme propensity scores via the overlap weights. *American Journal of Epidemiology*. 188(1), 250-257.

34. Ding P, and **Li F**.(2019). A bracketing relationship between difference-in-differences and lagged-dependent-variable adjustment. *Political Analysis*. 27(4), 605-615.

35. Li F*, **Li F**. (2019). Double-robust estimation in difference-in-differences with an application to traffic safety evaluation. *Observational Studies*. 5, 1-20.

36. Li F*, **Li F**. (2019). Propensity score weighting for causal inference with multiple treatments. *Annals of Applied Statistics*. 13(4), 2389-2415. (an earlier version won JSM 2019 Biometrics Section student paper award)

37. Dong J*, Zhang J, Zeng S*, and **Li F**. (2020). Subgroup balancing propensity score. *Statistical Methods in Medical Research*. 29(3) 659–676.

38. Lu D, Guo F, **Li F**. (2020). Evaluating the causal effects of cellphone distraction on crash risk using propensity score methods. *Accident Analysis and Prevention*. 143, 105579.

39. Thomas LE, **Li F**, Pencina M. (2020). Using propensity score methods to create target populations in observational clinical research. *Journal of American Medical Association*. 323(5):466-467.

40. Thomas LE, **Li F**, Pencina M. (2020). Overlap weighting: a propensity score method that mimics attributes of a randomized clinical trial. *Journal of American Medical Association*. 323(23):2417-2418.

41. Rosenbaum S, Zeng S*, Campos FA, Gesquiere LR, Altmann J, Alberts SC, **Li F**, Archie EA. (2020). Social bonds do not mediate the relationship between early adversity and adult glucocorticoids in wild baboons. *Proceedings of the National Academy of Sciences*. 33: 20052-20062

42. Zeng S*, **Li F**, Ding P. (2020). Is being an only child harmful to psychological health?: Evidence from an instrumental variable analysis of China's One-Child Policy. *Journal of Royal Statistical Society - Series A*. 183(4), 1615-1635.

43. Lu D, Tao C, Chen J, **Li F**, Guo F, Carin L. (2020). Reconsidering generative objectives for counterfactual reasoning. *34th Conference on Neural Information Processing Systems (NeurIPS2020)*.

44. Zhang YN, Chen Y, Wang Y, **Li F**, Pender M, Wang N, Yan F, Ying XH, Tang SL, Fu CW. (2020). Reduction in healthcare services during the COVID-19 epidemic in China. *BMJ Global Health*. 5:e003421. doi:10.1136/bmjgh-2020-003421.

45. Zeng S*, Li F, Wang R, **Li F**. (2021). Propensity score weighting for covariate adjustment in randomized clinical trials. *Statistics in Medicine*. 40(4), 842-858.

46. **Li F**, Mercatanti A, Mäkinen T, Silvestrini, A. (2021). A regression discontinuity design for ordinal running variable: Evaluating Central Bank purchases of corporate bonds. *Annals of Applied Statistics*. 15(1), 304-322.

47. Zeng S*, Rosenbaum S, Archie E, Alberts S, **Li F**. (2021). Causal mediation analysis for sparse and irregular longitudinal data. *Annals of Applied Statistics*. 15(2), 747-767.

48. Assaad S, Zeng S*, Tao C, Datta S, Mehta N, Henao R, **Li F**, Carin L. (2021). Counterfactual representation learning with balancing weights. *International Conference on Artificial Intelligence and Statistics 2021 (AISTAT)*. PMLR. 130: 1972-1980

49. Yang S*, Lorenzi E*, Papadogeorgou G*, Wojdyla D, **Li F**, Thomas LE. (2021). Propensity score weighting for causal subgroup analysis. *Statistics in Medicine*. 40:4294-4309. arXiv:2010.02121.

50. Yang S*, **Li F**, Thomas LE, Li F. (2021). Covariate adjustment in subgroup analyses of randomized clinical trials: A propensity score approach. *Clinical Trials*. 18(5). 570–581. (Finalist of Society of Clinical Trials (SCT) Thomas Chalmers Student Scholarship)

51. **Li F**, Tian Z, Bobb J, Papadogeorgou G, Li F. (2022). Clarifying selection bias in cluster randomized trials. *Clinical Trials*. 19(1), 33-41.

52. Zeng S*, **Li F**, Hu L, Li F. (2022). Propensity score weighting analysis for survival outcomes using pseudo observations. *Statistica Sinica*. Forthcoming. arXiv:2103.00605

53. Cheng C, **Li F**, Thomas LE, Li F. (2022). Addressing extreme propensity scores in estimating counterfactual survival functions via the overlap weights. *American Journal of Epidemiology*. 191(6), 1140-1151.

54. Wang Z*, Akande O, Poulos J*, **Li F**. (2022). Are deep learning models superior for missing data imputation in surveys?: Evidence from an empirical comparison. *Survey Methodology.* **48(2)**,375–399.

55. Zeng S*, Lange E, Campos F, Archie E, Alberts S, **Li F**. (2022). A Causal Mediation Model for Longitudinal Mediators and Survival Outcomes with an Application to Animal Behavior. *Journal of Biological, Environmental and Agricultural Statistics.* Forthcoming. arXiv:2104.08344.

56. Zhou T, Tong G, **Li F**, Thomas LE, Li F. (2022). PSweight: An R package for propensity score weighting analysis. *The R Journal.* 14(1):282-299.

57. Mäkinen T, **Li F**, Mercatanti A, Silvestrini, A. (2022). Causal analysis of central bank holdings of corporate bonds under interference. *Economic Modelling*. Forthcoming.

58. Papadogeorgou G*, Imai K, Lyall J, **Li F**. (2022) Causal inference with spatio-temporal data: Evaluating the effects of airstrikes on insurgent violence in Iraq. *Journal of Royal Statistical Society - Series B*. 84(5), 1969-1999. arXiv:2003.13555.

59. Li F, Tian Z, Tian Z, **Li F**. (2022). A note on identification of causal effects in cluster randomized trials with post-randomization selection bias. *Communications in Statistics – Theory and Methods*. Forthcoming.

60. Guo Q, Chen J, Wang D, Yang Y, Deng X, Carin L, **Li F**, Tao C*. (2022). Tight Mutual Information Estimation With Contrastive Fenchel-Legendre Optimization. *36th Conference on Neural Information Processing Systems (NeurIPS2022)*. arXiv:2107.01131

61. **Li F**, Ding P, Mealli F. (2023). Bayesian causal inference: a critical review. *Philosophical Transactions of the Royal Society A*. 381: 2022.0153.

62. Lange E, Zeng S*, Campos F, **Li F**, Tung J, Archie E, Alberts S. (2023). Early life adversity and adult social relationships have independent effects on survival in a wild animal model of aging. *Science Advances*. 9, eade717.

63. **Li F**, and Li F. (2023). Using propensity scores for racial disparities. *Observational Studies*. 9(1), 59-68.

Book Chapter

64. Zhang T, Sheng H, and **Li F**. (2016). Linear and Nonlinear Models for fMRI Time Series Analysis. *Handbook of Modern Statistical Methods: Neuroimaging Data Analysis*, Ombao H, Johnson W, Lindquist M, Aston J eds. Chapman and Hall - CRC Press.

65. **Li F**. (2022). Overlap weighting. *Handbook of Matching and Weighting Adjustments in Causal Inference*, J Zubizarreta, EA Stuart, D Small, PR Rosenbaum, eds. Chapman and Hall - CRC Press.

Discussions

66. Mealli F, and **Li F**. (2011). Discussion of "Transparent parametrization of models for potential outcomes" by Richardson, Evans and Robins. *Bayesian Statistics 9* (JM Bernardo, MJ Bayarri, JO Berger, AP Dawid, D Heckerman, AFM. Smith and M West eds.). Oxford University Press.

67. Papadogeorgou G*, and **Li F**. (2019). Discussion of "Penalized spline of propensity methods for treatment comparison" by Zhou, Elliot and Little. *Journal of the American Statistical Association*. 114(525):32-35.

68. Papadogeorgou G*, and **Li F**. (2020) Discussion of "Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects" by Hahn, Murray and Carvalho. *Bayesian Analysis*. 15(3): 1007-1013.

Preprints

69. **Li F**, Yu Y, Rubin DB. (2012). Imputing missing data by fully conditional models: Some cautionary examples and guidelines. *Duke University Department of Statistical Science Discussion Paper 11-24*.

70. Zeng S*, Assaad S, Tao C, Carin L, **Li F**. (2021). Double-robust representation learning for causal inference. arXiv:2010.07866.

71. Chen J, Gan Z, et al., **Li F**, Carin L, Tao C*. (2021) Simpler, Faster, Stronger: Breaking The log-K Curse On Contrastive Learners With FlatNCE. arXiv:2107.01152.

72. Yang S*, Zhou R*, **Li F**, Thomas LE. (2023). Propensity Score Methods for Causal Subgroup Analysis with Time-to-Event Outcomes.

73. Liu B*, Wruck L, **Li F**. (2022). Principal stratification for noncompliance with time-to-event outcomes. arXiv:2301.07672

74. Chang* C-R, Song Y, **Li F**, Wang R. (2022). Covariate adjustment in randomized experiments with incomplete covariate and outcome data.

75. Cheng C, Guo G*, Liu B, Wruck L, **Li F**, F Li. (2023). Multiply robust estimation for causal survival analysis with treatment noncompliance. arXiv:2305.13443.

SOFTWARE PACKAGE

1. `PSweight` (2020): Propensity Score Weighting for Causal Inference. Tianhui Zhou, Guangyu Tong, Fan Li, Laine Thomas, Fan Li. `https://CRAN.R-project.org/package=PSweight`

2. `PStrata` (2022): Principal Stratification for Causal Inference. Bo Liu, Fan Li. `https://CRAN.R-project.org/package=PStrata`

GRANTS

1. Innovative Biostatistical Methods for Analysis and Assessment of Clinical Trials Augmented by Real World Data. Burroughs Wellcome Fund Innovation in Regulatory Sciences Award. 2021-2026. Role: Co-PI (PI: Laine Thomas). Total cost: $500,000.
2. COVID-19 Enhancement: Methods for the Design and Conduct of Subgroup Analysis in Observational Studies. PCORI ME-2018C2-13289, 2019-2023. Role: Co-I (PI: Laine Thomas). Total cost: $349,999.
3. New causal inference methods for cluster randomized trials with post-randomization selection-bias. PCORI ME-2019C1-16146, 2020-2023. Role: PI. Total cost: $946,222
4. Methods for the design and conduct of subgroup analysis in observational studies. PCORI ME-2018C2-13289, 2019-2022. Role: Co-I (PI: Laine Thomas). Total cost: $731,268
5. The biodemography of early adversity: social behavioral processes in a wild animal model. NIH 1R01 AG053308-01A1, 2018-2023. Role: Co-PI (PI: Susan Alberts). Direct cost: $1,542,592
6. A life course perspective on the effects of cumulative early adversity on health. NIH 1R01 AG053330-01A1, 2017-2022. Role: Co-PI (PI: Beth Archie). Total cost: $2,352,291
7. Religion, Spirituality and CVD Risks: A Focus on African Americans. NIH 5R01MD011606-02, 2017-2022. Role: Statistical Investigator (PI: Bentley-Edwards). Total cost: $2,831,644
8. Prospective Multicenter Observational Cohort Study of Comparative Effectiveness of Disease-Modifying Treatments for Myasthenia Gravis (MG). PCORI R-1609-35953, 2017-2020. Role: Statistical Investigator. (PI: Don Sanders). Total cost: $2,517,289
9. New weighting methods for causal inference. NSF-SES 1424688, 2014-2017. Role: PI. Total cost: $190,000.
10. Bayesian multivariate analysis for causal inference with intermediate variables. NSF-SES 1155697, 2012-2015. Role: PI. Total cost: $80,000.
11. Collaborative research: Statistical modeling and inference for high-dimensional multi-subject neuroimaging data. NSF-DMS 1208983, 2012-2015. Role: PI. Total cost: $71,100.
12. The Triangle Census Research Network. NSF-NCRN, 2011-2016. Role: Investigator (PI: Jerry Reiter).

MENTORING

Doctoral Advisees

Scott Schwartz	2010 Statistical Geneticist and Bioinformatics Scientist, Texas A&M University
Nghi Maggie Nguyen	2018 Research Scientist, Duke University Department of Neurology
Fan (Frank) Li	2019 (Biostatistics& Bioinformatics) Assistant Professor, Yale University Department of Biostatistics

Abbas Zaidi	2019 (co-advise with Sayan Mukerjee) AI researcher, Facebook
Elizabeth Lorenzi	2019 Statistical Scientist, Berry Consultants
Shuxi Zeng	2021 Research Scientist, Facebook
Siyun Yang	2022 (co-advise with Laine Thomas, B& B) Research Scientist, Facebook
Bo Liu	2021-
Yueqi Guo	2022-

Postdoctoral Mentees

Georgia Papadogeorgou	2018-2020 (co-advise with David Dunson) Assistant Professor, University of Florida Department of Statistics
Jason Poulos	2019-2021 Postdoctoral Fellow, Harvard Medical School Department of Health Care Policy
Chenyang Tao	2021 Applied Scientist, Amazon
Ruiwen Zhou	2021-2022 (co-advise with Laine Thomas)

Master Advisees

Ying Yang (Neurobiology, MS)	2011
Olanrewaju Akande (Statistical Science, MSEM)	2015
Eve Oh (Statistical Science MSEM)	2015
Shuo Wang (MSS), Joon Sup Park (MSS) Robert Wan (MIDS), Chengxin Yang (MSS)	2022

Undergraduate advisees

Colin Hwang	2011
Ekaterina Petrova	2012
Jack Fu	2013
Tracy Qi Dong	2014
Fiamma Li	2015
Anna Jiang	2016
Jerry Chia-Rui Chang	2019
Pei Yi Zhuo	2023

Doctoral thesis committee

2011 Hongxia Yang, Chiranjit Mukherjee
2012 Yajuan Si, Jochi Nakajima, Kai Cui
2013 Fangpo Wang, Jared Murray
2015 Monika Jincheng Hu, Tsuyoshi Kunihama
2016 Tracy Schifeling, Feifei Wang (Peking University)
2018 Victor Pena
2019 Olanrewaju Akande, Jodi Heck Wortman, Phil White
2020 Danni Lu (Virginia Tech)

Preliminary oral committee

2009 Hongxia Yang, Chiranjit Mukherjee, Minhui Shi
2010 Fangpo Wang, Yajuan Si, Jochi Nakajima
2011 Kai Cui
2012 Tsuyoshi Kunihama
2014 Michael Lindon
2015 Victor Pena
2016 Jody Heck Wortman, Elizabeth Lorenzi
2017 Kyle Burris, Abbas Zaidi, Olanrewaju Akande, Phil White
2019 Shuxi Zeng
2021 Serge Assaad

Master thesis committee

2010 Shouqiang Wang (Operational Research), Arturas Rozenas (Pol Sci)
2012 Yiting Deng (Computer Science)
2014 Yingjian Wang (ECE)
2019 Gauri Kamat, Yunji Zhou (B&B)
2020 Yangfan Ren
2021 Haoling Zheng, Marco Morucci (Pol Sci)
2022 Yi Liu (B&B)

Undergraduate thesis committee

2018 Andrew Cooper
2019 Vivek Sriram
2020 Daniel Spottiswood

TEACHING

(All in Department of Statistical Science, Duke University)

STA 130 Probability and Statistics in Engineering (2010F, 2012-14F, 2012S, 2015S)
STA 320 Design and Analysis of Causal Studies (2011F, 2014S, 2016S)
STA 440 Case Studies in the Practice of Statistics (2019F)

STA 610 Hierarchical models (2023F)
STA 611 Introduction to Mathematical Statistics (2008F)
STA 640 Causal Inference (2015F, 2017-18F, 2020F, 2021-2023S)
STA 723 Statistics Case Studies (2014-19S)
STA 732 Statistical Inference (2009-10S)
STA 790 Special Topics: Causal Inference (2009F), Bayesian Causal Inference (2022F)

PROFESSIONAL APPOINTMENTS AND SERVICE

Editorial Boards

2023-	Editor for Social Science, Biostatistics and Policy, *Annals of Applied Statistics*
2023-24	Guest Editor, Special Issue on "Causal Inference: past, present, and future" *The New England Journal of Statistics in Data Science (NEJSDS)*
2016-2023	Associate Editor, *Bayesian Analysis*
2019-	Associate Editor, *Observational Studies*
2020-	Associate Editor, *Journal of American Statistical Association - TM*
2016-2019	Associate Editor, *Journal of American Statistical Association - ACS*
2013-2017	Associate Editor, *Journal of Statistical Theory and Practice*
2018	Associate Editor, *The American Statistician* special issue on "Statistical inference in the 21th century"

Peer Review Activities

American Statistician, Annals of Applied Statistics, Annals of Internal Medicine, Bayesian Analysis, Biostatistics, Biometrics, Biometrika, BMC Research Methodology, BMJ, Canadian Journal of Statistics, Circulation, Computational Statistics and Data Analysis, Health Services and Outcomes Research Methodology, International Journal of Methods in Psychiatric Research, Journal of Causal Inference, Journal of Computational and Graphical Statistics, JAMA, JAMA Cardiology, JAMA Network Open, Journal of American Statistical Association, Journal of Applied Econometrics, Journal of Causal Inference, Journal of Royal Statistical Society (Series A, B, C), Journal of Statistical Planning and Inference, Neuroimage, Observational Studies, Psychometrika, Scandinavian Journal of Statistics, Statistical Methods in Medical Research, Statistica Sinica, Statistical Science, Statistics and Computing, Statistics in Medicine, Statistics and Probability Letters, Survey Methodology.

Grant Review Panel

National Science Foundation	2013, 2015, 2016, 2018
National Health Institute - BMRD	2016

Ad-hoc Review of Grant Proposals

Netherlands Organisation for Scientific Research (NWO)
Natural Sciences and Engineering Research Council of Canada (NSERC)
Canadian Statistical Sciences Institute (CANSSI)
Health Effects Institute

Conference and Workshop Organizing

2013-14 Group leader, Causal Inference working group, SAMSI CMSS program
2015 Organizer, the G70 Conference: A Celebration of Alan Gelfand's 70th Birthday, Durham
2017 Organizer, NISS workshop on causal inference and machine learning/high dimensional statistics at Atlantic Causal Inference Conference (ACIC), UNC-Chapel Hill
2018 IMS Program Chair, ENAR spring meeting, Atlanta
2019 Organizer, Bayesian causal inference workshop, MBI, Ohio State University
2019 Organizer, Opening workshop of SAMSI Causal Inference Program, Duke University
2020 Organizer, SAMSI Causal Inference Program
2021-22 Member, ISBA 2022 World Meeting Program Committee

Professional Societies

2018, 20 Member, Nominating Committee, International Society for Bayesian Analysis (ISBA)
2019 Member, Selecting Committee for the founding co-editors of the IMS Data Science Journal
2022 Member, Mitchell Prize Selection Committee, ISBA
2023-2024 Member, Committee on Nominations, Institute of Mathematical Statistics

Promotion and External Reviews

2019- Promotion review (Yale, Peking, U Wisconsin at Madison, U Michigan)
2022 Member of External Review Panel of Department of Statistics and Data Science, Wharton School of Business, University of Pennsylvania

ACADEMIC SERVICE

Department of Statistical Science

2009-10, 17 First Year PhD Exam Coordinator
2009-16, 19-20 PhD Admissions Committee
2010-12 Seminar Series Coordinator
2013, 16-17 Master's Program Admissions Committee
Master's Program Director
2017- Master's Program Advisory Committee
2018, 22 Tenure-Track Faculty Search Committee
2019 PhD Program Evaluation committee
2021 DST faculty search committee chair

Duke University

2014 Faculty compensation equity committee
2014-16 Academic Council
2017-22 Academic Program Committee (APC)
2018-19 Search Committee for Chair of Department of Biostatistics & Bioinformatics
2019-20 Search Committee for Executive Vice Chancellor at Duke Kunshan University
2020-21 Duke Strategy Team 2030 Faculty Group
2021 Duke 2030 Working Group on Research
2021-2023 Duke Kunshan University (DKU) Faculty Hearing Committee
2022 Review Committee of the Executive Vice Provost

PRESENTATIONS

Short Course and Tutorial

1. (2011) Short course on "Statistical Methods in Causal Inference". Finnish Society of Epidemiology. Helsinki, Finland.
2. (2017) Tutorial on propensity score methods in traffic safety research. Transportation Research Board Annual Meeting. Washington, DC.
3. (2017) Short course on "New weighting methods in comparative effectiveness research". Duke-Industry Statistics Symposium 2017. Durham, NC.
4. (2018) Tutorial on "Causal inference". Duke Plus Data Science, Durham, NC.
5. (2019) Short course on "Bayesian causal inference". Atlantic Causal Inference Conference, Montreal, Canada.
6. (2019) Tutorial on "Bayesian causal inference". Bayesian Causal Inference Workshop, Ohio State University, Columbus, OH.
7. (2020) Tutorial on "New weighting methods for comparative effectiveness research." International Conference on Health Policy Statistics 2020, San Diego, CA.
8. (2023) Tutorial on "Propensity score weighting for comparative effectiveness research: methods, new developments and software". International Conference on Health Policy Statistics 2023, Scottsdale, AZ.
9. (2023) Short course on "Bayesian causal inference". Applied Bayesian Summer School 2023, Florence, Italy.
10. (2023) Short course on "Causal inference". Columbia University, Department of Statistics.

Seminars

1. (2023) McGill University, Department of Epidemiology, Biostatistics and Occupational Health, Keynote speaker at Student Career Day
2. (2023) University of Cambridge, MRC Biostatistics Unit (virtual)
3. (2023) University of Michigan, Department of Statistics
4. (2022) Texas A&M University, Department of Statistics
5. (2022) Georgia Tech ISyE Statistics Seminars
6. (2022) DCRI Clinical Research Fellowship Program
7. (2022) Duke University Department of Philosophy Causation Group
8. (2022) Michigan State University Department of Statistics and Probability (virtual)
9. (2022) Online Causal Inference Seminar (OCIS) Series (virtual)
10. (2022) OHDSI Methods Working Group, UCLA
11. (2022) International Biometric Society Journal Club

12. (2022) Criteo AI lab (virtual)
13. (2021) Online interdisciplinary seminars on statistical methodology for social and behavioral research, University of Connecticut (virtual)
14. (2021) Duke University, Department of Population Health Sciences (virtual)
15. (2021) Harvard School of Public Health, Working Group on Causal Inference and Machine Learning (virtual)
16. (2021) University of Pennsylvania, Center for Causal Inference (virtual)
17. (2021) Carnegie Mellon University, Department of Statistics and Data Science (virtual)
18. (2021) Online Causal Inference Seminar (OCIS) Series (virtual)
19. (2020) Icahn School of Medicine at Mount Sinai, Institute for Translational Epidemiology (virtual)
20. (2020) University College London, Department of Statistical Science (virtual)
21. (2020) Duke University, Plus Data Science, COVID-19 Data Science Seminar (virtual)
22. (2020) University of Chicago, Department of Statistics
23. (2020) Vanderbilt University, Department of Biostatistics
24. (2019) University of Michigan, Department of Biostatistics
25. (2019) Brown University, Department of Biostatistics
26. (2019) University of Pennsylvania, Department of Statistics
27. (2018) University of Pennsylvania, Department of Biostatistics, Epidemiology and Informatics
28. (2018) Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics
29. (2018) North Carolina State University, Department of Statistics
30. (2018) University of Texas School of Public Health, Department of Biostatistics and Data Science
31. (2018) SAS, Cary, NC
32. (2017) Virginia Tech, Department of Statistics
33. (2016) University of California, Berkeley, Department of Statistics, Neyman Seminar
34. (2016) Duke University, Comparative Effectiveness Research Program
35. (2016) Duke Clinical Research Institute, Duke University
36. (2016) University of Maryland at Baltimore, Department of Mathematics
37. (2015) Tsinghua University (China), Center for Statistical Science
38. (2015) University of Turku (Finland), Department of Mathematics
39. (2015) University of North Carolina at Chapel Hill, Causal inference research group
40. (2014) University of North Carolina at Chapel Hill, Department of Biostatistics
41. (2013) Durham Veterans Administration, Division of Health Services Research and Development

42. (2013) Cornell University, Weill Medical College, Department of Public Health, Division of Biostatistics and Epidemiology
43. (2013) Collegio Carlo Alberto, University of Turin, Italy
44. (2012) University of Florence, Department of Statistics, Italy
45. (2012) University of North Carolina at Chapel Hill, Center for Developmental Science
46. (2012) Ohio State University, Department of Statistics
47. (2012) IBM Watson Research Center
48. (2012) Columbia University, Department of Psychiatry, Division of Biostatistics
49. (2011) University of Pennsylvania, Department of Statistics
50. (2011) University of North Carolina at Chapel Hill, Causal inference research group
51. (2011) University of Virginia, Department of Statistics
52. (2011) Brown University, Center for Statistical Sciences
53. (2008) Duke University, Department of Statistical Science
54. (2008) University of Maryland at College Park, Department of Epidemiology and Biostatistics
55. (2008) University of North Carolina at Chapel Hill, Department of Biostatistics
56. (2007) Fox Chase Cancer Center, Biostatistics Facility
57. (2006) Harvard University, Department of Health Care Policy
58. (2006) Group Health Cooperative, Center of Health Studies
59. (2006) University of Chicago, Department of Health Studies
60. (2006) University of Pittsburgh, Department of Statistics
61. (2006) Ohio State University, Department of Statistics

Invited Conference Presentations

1. (2023) ENAR Spring Meeting, Nashville, TN
2. (2022) BAYES2022 - Bayesian Biostatistics Conference, Bethesda, MD
3. (2022) JSM 2022, Washington DC
4. (2022) ISBA World Meeting, 2022, Montreal, Canada
5. (2022) Workshop on Complex Data with Missingness, Measurement Errors, and High Dimensionality, Banff International Research Station (virtual)
6. (2021) Workshop on Computational Advertising, Banff International Research Station (virtual)
7. (2021) Pacific Causal Inference Conference (PCIC) 2021 (virtual)
8. (2021) JSM 2021 (virtual)
9. (2021) ISBA 2021 World Meeting (virtual)

10. (2021) SAMSI Opening Workshop on Data Science in the Social and Behavioral Sciences (virtual)
11. (2020) SAMSI Games, Decisions, Risk and Reliability (GDRR) Program Transportation Workshop, Durham, NC
12. (2019) Translating Duke Health Immunology & Transplant Initiative Symposium, Duke University, Durham.
13. (2019) JSM, Denver, CO
14. (2019) ICSA China Conference, Tianjin, China
15. (2019) Atlantic Causal Inference Conference 2019, Montreal, Canada
16. (2019) ENAR Spring Meeting, Philadelphia, PA
17. (2019) University of Florida, Gainsville. UF Winter Statistics Workshop.
18. (2018) JSM, Vancouver, Canada
19. (2018) Conference on Evidence and the Individual Patient: Understanding Heterogeneous Treatment Effects for Patient-Centered Care. National Academy of Medicine, Washington, DC
20. (2018) Webinar, Predictive Analytics and Comparative Effectiveness (PACE) Center, Tufts Medical Center.
21. (2018) ENAR Spring Meeting. Atlanta, GA
22. (2017) International Workshop on Objective Bayes Methodology (O-Bayes17). Austin, TX
23. (2017) SAMSI summer workshop on transportation statistics, Durham, NC
24. (2017) Joint Statistical Meeting, Baltimore, MA
25. (2017) European Meeting of Statisticians, Helsinki, Finland
26. (2017) Atlantic Causal Inference Conference 2017, UNC-Chapel Hill
27. (2016) University of Columbia, Department of Statistics, Causal Inference Conference
28. (2016) Fourth International Conference on the Interface between Statistics and Engineering, Palermo, Italy
29. (2016) ISBA 2016 World Meeting, Sardinia, Italy
30. (2016) Atlantic Causal Inference Conference, New York City
31. (2016) Technical Advisory Committee (TAC) annual meeting, Federal Highway Adminstration, McLean, Virginia
32. (2014) SAMSI Computational Methods in Social Sciences Program Transition Workshop, Durham, NC
33. (2014) ENAR spring meeting, Baltimore, MD
34. (2013) Technical experts meeting on statistical methodologies, Federal Highway Adminstration (FHWA), Durham, NC
35. (2013) International Workshop on Objective Bayes Methodology, Durham, NC
36. (2013) Joint Statistical Meeting, Montreal, Canada

37. (2013) ENAR spring meeting, Orlando, FL
38. (2013) SAMSI Computational Methods in Social Sciences Program Opening Workshop, Durham, NC
39. (2013) SAMSI Neuroimaging Data Analysis Summer Program, Durham, NC
40. (2012) ISBA 2012 World Meeting, Kyoto, Japan
41. (2012) ENAR spring meeting, Washington, DC
42. (2012) 5th Annual Bayesian Biostatistics Conference, Houston, TX
43. (2011) Joint Statistical Meeting, Miami, FL
44. (2011) IISA Conference on Probability, Statistics, and Data Analysis, Raleigh, NC
45. (2010) The Eighth ICSA International Conference, Guangzhou, China
46. (2007) Joint Statistical Meeting, Salt Lake City, UT

WPATH Assertion # 1

There is strong evidence demonstrating the benefits in quality of life and well-being of gender-affirming treatments, including endocrine and surgical procedures, properly indicated and performed as outlined by the Standards of Care (Version 8), in TGD people in need of these treatments (e.g., Ainsworth & Spiegel, 2010; Aires et al., 2020; Aldridge et al., 2020; Almazan & Keuroghlian, 2021; Al-Tamimi et al., 2019; Balakrishnan et al., 2020; Baker et al., 2021; Buncamper et al., 2016; Cardoso da Silva et al., 2016; Eftekhar Ardebili, 2020; Javier et al., 2022; Lindqvist et al., 2017; Mullins et al., 2021; Nobili et al., 2018; Owen-Smith et al., 2018; Özkan et al., 2018; T'Sjoen et al., 2019; van de Grift, Elaut et al., 2018; White Hughto & Reisner, Poteat et al., 2016; Wierckx, van Caenegem et al., 2014; Yang, Zhao et al., 2016). Page S18.

Name	Study Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
van de Grift, Elaut et al., 2018 DOI: 10.1080/0092623X.2017.1326190	Prospective, no before-after for most outcomes	Multicenter, cross-sectional prospective study (follow up time 4-6 years) of 201 persons diagnosed with gender dysphoria who applied for medical interventions from 2007 until 2009. Main outcome measures were procedure performed, self-reported complications, and satisfaction with surgical outcomes (standardized questionnaires), QoL (Satisfaction With Life Scale, Subjective Happiness Scale, Cantril Ladder), gender dysphoria (Utrecht Gender Dysphoria Scale), and psychological symptoms (Symptom Checklist-90). The majority of trans women had undergone a vaginoplasty, and some also received mamma augmentation.	Assesse the outcomes of **gender-affirming surgery (GAS)** 4 to 6 years after first clinical contact, and the associations between postoperative (dis)satisfaction and quality of life (QoL).	Postoperative satisfaction was 94% to 100%, depending on the type of surgery performed. Only a few study participants reported feelings of regret. Satisfied respondents'QoL scores were similar to reference values; dissatisfied or regretful respondents' scores were lower. Psychological symptoms and life dissatisfaction at baseline were associated with treatment dissatisfaction at follow-up.		1. Only the SymptomChecklist 90-R (SCL-90) measure of gender dysphoria is measured both before and after SRS, all other outcomes (e.g. QoL) are only measured after SRS. 2. Selection bias due to low response rate: 201 out of 546 (37%) eligible subjects responded. The paper provides a justification that the respondent and non respondents baseline characteristics are similar, but provided no data. 3. Missing data: 20% to 67% missing data in all but one self-reported outcomes.	Despite being a prospective study, this paper has several methdological shortcomings, as outlined in the limitations section. It focuses on comparing the dissatified and satistified groups, and comparing to reference groups, but **didn't conduct direct before-after comparison** of the same patients (e.g. paired t-test). This paper finds that psychological symptoms and life dissatisfaction at baseline were significantly associated with treatment dissatisfaction at follow-up. This shows the importance of adjusting for the baseline mental status, which is available in most studies in this field.
Cardoso da Silva et al., 2016 DOI: 10.1016/j.jsxm.2016.03.370	Prospective, before-after	Prospective study of 47 Brazilian male-to-female transsexual individuals. QoL is measured using the WHOQOL-100 (100-item World Health Organization Quality of Life Assessment) and sociodemographic questions. Initial assessment and 1 year after SRS (sex reassignment surgery).	Assess the impact of **surgical interventions** on quality of life (QoL) of male-to-female transsexual individuals.	The participants showed significant improvement after SRS in domains II (psychological) and IV (social relationships) of the WHOQOL-100. In contrast, **domains I (physical health) and III (level of independence) were significantly worse after SRS.** Individuals who underwent additional surgery had a decrease in quality of life reflected in domains II and IV one year after SRS. Domains for the environment and spirituality, religion, and personal beliefs domains did not change after SRS	Prospective study, with pre- and post-measure of the same patients		This is a prospective studies which provide pre- and post- surgery comparison of the same patients. And **the results are mixed**. In particular, it found "domains I (physical health) and III (level of independence) were significantly worse after SRS"

Lindqvist et al., 2017 DOI: 10.1007/s00238-016-1252-0	Prospective cohort, before-after, longitudinal	A prospective cohort study on 190 patients undergoing male-to-female GRS at Karolinska University Hospital between 2003 and 2015. SF-36 QoL score pre-operatively, 1, 3, and 5 years post-operatively.	examine the QoL of transgender women undergoing **gender reassignment surgery (GRS)**.	1. SF-36 scores are higher 1 year post-GRS compared to pre-operatively (but the statistical significance is on boundary, <0.05), and are lower 5 years post-GRS compared to pre-operatively. There is a strong statistical significant trend of decline health trendy post 1 year. 2. On most dimensions of SF-36, transgender women reported a lower QoL than the general population.	Prospective cohort, with data on both pre- and post-operation at multiple time points	the statistical significance declared on the increase in QoL in year 1 compared to year 0 is only at p=0.05 level, while the later decrease is much more stronger level (p=0.0001)	The paper in fact **shows that 3 and 5 years post-operation the QoL is worse than pre-operation**, but didn't report the associated stat significance of the comparison between year 3/5 and year 0 among transgender women; neither did they report that of the comparision in general population. Only vaguely commented that the downward trend is similar between transgender women and general population, but again without any accurate measure of similarity or the stat significance in the comparisons in general population
Aires et al., 2020 DOI: 10.1080/26895269.2020.1848690	Prospective, before-after	**Prospective cohort** of transgender women submitted to **chondrolaryngoplasty** between March 2018 and October 2019. Two before and after measurements: (1) voice analysis by therapist (GRBAS score), before and 1-month after; (2) visual analog scale for aesthetic satisfaction, before and 6-month after. Sample size: 15	To assess and compare acoustic and perceptual voice outcomes and aesthetic satisfaction of transgender women submitted to **chondrolaryngoplasty (or "tracheal shaving", cosmetic surgery to reduce the laryngeal prominence)**	1. No difference in GRBAS score pre- and post- surgery. 2. Significant improvement in the visual analog scale for aesthetic satisfaction.	Prospective	Small sample size (15); short follow-up period (6 months)	This paper focuses on a special type of surgery of chondrolaryngoplasty or tracheal shaving. The outcomes are specific to voice and visual aesthetic satisfaction; there is no information on mental outcomes such as QoL, gender dysphoria or general physical outcomes.
Wierckx, K., Van Caenegem, et al. (2014). Journal of Sexual Medicine, https://doi.org/10.1111/jsm.12571	Prospective, before-after	Multicenter 1-year prospective study in 53 trans men and 53 trans women.	Report the short-term effects of **cross-sex hormone therapy** (CHT) on hormonal and clinical changes, side effects, and adverse events in trans men (female-to-male gender dysphoric persons) and trans women (male-to-female	Current treatment modalities were effective and carried a low risk for side effects and adverse events at short-time follow-up	Prospective cohort study with short follow up, with data on both pre- and post-treatment data		This paper focuses on the effect of **hormone therapy**, specifically on safety, not quality of life. It stated "Data on the effects of cross-sex hormone therapy (CHT) are limited due to the low prevalence of gender dysphoria, small number of subjects treated at each center, lack of prospective studies, and wide variations in treatment modalities."

Ainsworth & Spiegel, 2010 DOI: 10.1007/s11136-010-9668-7	Retrospective cross-sectional	**Facial Feminization Surgery** outcomes evaluation survey and the SF-36 quality of life survey were administered to male-to-female transgender individuals via the Internet and on paper. Sample size is 247.	To determine the self-reported quality of life of male-to-female (MTF) transgendered individuals and how this quality of life is influenced by **facial feminization and gender reassignment surgery.**	(1) Mental health-related QoL was statistically lower (P<0.05) in transgendered women without surgical intervention compared to the general female population and transwomen who had gender reassignment surgery (GRS), facial feminization surgery (FFS), or both. (2) no statistically significant difference in the mental health-related quality of life among transgendered women who had GRS, FFS, or both. (3) Participants who had FFS scored statistically higher (P<0.01) than those who did not in the FFS outcomes evaluation		Retrospective study. **No pre-post comparison of the same participate**; subject to confounding.	This paper focuses on QoL of MTF transwomen following **facial feminization surgery**. It is a retrospective study, no before-after comparison of the same participant
Almazan & Keuroghlian, 2021DOI: 10.1001/jamasurg.2021.0952)	Retrospective cross-sectional	**Secondary analysis the 2015 US Transgender Survey** (27715 respondents total). Two comparison groups: 3559 (12.8%) underwent 1 or more types of GAS at least 2 years prior to survey, 16401 (59.2%) endorsed a desire for GAS but denied undergoing any of these.	Evaluate **associations between gender-affirming surgeries (GAS)** and mental health outcomes	Statistically significant association between gender-affirming surgery and improved mental health outcomes (past-month psychological distress, past-year smoking, past-year suicidal ideation), all at 0.001 level. But no significant association between with GAS and past-month binge alcohol use.	Large sample size (largest existing data set containing comprehensive information on the surgical and mental health experiences of TGD people)	(1) Restropective study: no pre- and post- information of the same subject. Lack the single most important predictor: baseline mental health status. **The paper gave a post hoc justification:** "Our post hoc analysis demonstrates that lifetime suicidality and substance use behaviors are not associated with the exposure variable in this sample. Therefore, prior mental health factors do not appear to confound associations between gender-affirming surgery and subsequent mental health outcomes in our study." **The "justification" is statistically unfounded and deeply flawed, has no scientific merit**. (2) GAS was associated with lower odds of past-year suicidal ideation, but no statistically significant association between GAS and past-year suicide attempts; (3) (1) **Association study, no causal inference adjustment**	It is a large restrospective study, but lacks pre- and post- surgery information of the same patients. Therefore, the analysis did not adjust for the confounding bias due to the difference in the baseline outcome -- the most important confounder.

Al-Tamimi et al., 2019 DOI: 10.1016/j.jsxm.2019.07.027	Retrospective cross-sectional	Retrospective descriptive study of 83 transgender men who went through **secondary phalloplasty (a type of genital GAS**). Main outcome measures are surgical techniques, patient motivation, and outcomes of secondary phalloplasty after metoidioplasty in transgender men.	Explore the reasons for secondary phalloplasty, describe the surgical techniques, and report on the clinical outcomes.	A wide variety of surgical techniques were used to perform secondary phalloplasty. Several types of complications are reported on some patients. (Intraoperative complications (revision of microvascular anastomosis) occurred in 3 patients (5.5%) undergoing free flap phalloplasty. Total flap failure occurred in 1 patient (1.2%). Urethral fistulas occurred in 23 patients (30.3%) and strictures in 27 patients (35.6%))		Retrospective study	This is a descriptive study, describing the surgical techniques, and reporting on the clinical outcomes. **It is not a comparative study on the effect of the surgery. It did not report QoL or satisfaction.**
Mullins et al. 2021: DOI: 10.1542/peds.2020-023549	Retrospective	A retrospective chart review was conducted at a pediatric hospital-associated transgender health clinic. The primary outcome was incidence of arterial or venous thrombosis during gender-affirming hormone therapy (GAHT). Secondary measures included the prevalence of thrombosis risk factors.	To examine thrombosis and thrombosis risk factors among an exclusively **adolescent and young adult transgender population.**	These data suggest that GAHT in youth, titrated within physiologic range, does not carry a significant risk of thrombosis in the short-term, even with the presence of preexisting thrombosis risk factors.		Retrospective	This paper examines thrombosis and thrombosis risk factors among adolescent and young adult transgender population who received GAHT. The paper **does not provide any information in the effects of GAHT on quality of life and well-being.**
Balakrishnan et al., 2020 DOI: 10.1097/PRS.0000000000002684	Retrospective	A retrospective study of 42 transwomen performed from 2007 to 2017 in one hospital, followed up for 45 months. A validated institutional score for subjective assessment and objective assessment were used at the end of follow-up period.	Assess the esthetic outcomes of PAM (**Augmentation Mammoplasty**) - a type of surgery.	92.85% of transwomen achieved grade-A score with both subjective and objective assessment scoring system.		retrospective, descriptive (not comparative)	This study is a **descriptive analysis** of the outcomes of PAM; it is **not a comparative effectiveness analysis: doesn't compare QoL between pre- and post- PAM; or between PAM and non-PAM subjects.**
Buncamper et al., 2016 DOI: 10.1055/s-0040-1709427	Retrospective	Retrospective study of 475 patients who underwent penile inversion vaginoplasty	Assess intraoperative and postoperative complications after **penile inversion vaginoplasty**	Successful vaginal construction without the need for secondary functional reoperations was achieved in the majority of patients. Intraoperative complications are scarce. Postoperative complications occur frequently but are generally minor and easily treated.		retrospective, and descriptive (not comparative)	This is a **descriptive analysis** of the clinical outcomes (e.g. complications) of penile inversion vaginoplasty. There is no pre- vs. post- surgery or surgery vs. non-surgery comparisons. No discussion about the necessity of the surgery

Owen-Smith et al., 2018 DOI: 10.1016/j.jsxm.2018.01.017	Retrospective cross-sectional	Cross-sectional comparison of self-reported outcomes between patients who had GCT vs. not. A cohort of transgender individuals recruited from 3 health plans in Georgia, Northern California, and Southern California who completed a survey. Outcomes of interest included body-gender congruence, body image satisfaction, depression, and anxiety.	Examine the association between gender confirmation treatments (GCT) and individuals' body-gender congruence, body image satisfaction, depression, and anxiety in cohorts of transmasculine (TM) and transfeminine (TF) individuals.	The proportion of participants with low body-gender congruence scores was significantly higher in the "no treatment" group compared to the definitive bottom surgery group (PR 3.96). The PR for depression comparing participants who reported no treatment relative to those who had definitive surgery was 1.94; the corresponding PR for anxiety was 4.33.) No significant difference in any other comparisons (outcome and surgery types).	Sample include both transman and transwoman	1. Retrospective cross-sectional. No before- and after- surgery comparison. 2. May not be representative of the transgender population in the United States because (i) the study sample is collected from only three health systems in GA and CA; (i) Low response rate (33% 697 out of 2136). 3. **No comparison between non GCT and surgery types other than definitive bottom surgery**. 4. Reporting is **cherry picking**: many comparisons (different outcomes (e.g. body image), different surgery groups), only reported the ones which are statistically significant (3 out of 10 comparisons)	Combination of several treatments, including hormone therapy (HT) and/or surgical change of the chest and genitalia ("top" and "bottom" gender confirmation surgeries). The results are mixed, but the reporting focuses on the subset of statistically signficant ones.
Yang, Zhao et al., 2016 DOI: 10.1016/j.jsxm.2016.03.369	Retrospective cross-sectional	Cross-sectional analysis of 247 transwomen in Shenyang, China. Self reported QOL. Only 4 underwent SRS	To assess QOL (both physical and mental) of trans women in Shenyang, China and associated factors	Chinese transgender women reported high levels of physical QOL but low levels of mental QOL. Their mental QOL was more pronounced than their physical QOL. Transition status and sexual partnership played the most important roles in physical and mental health. Furthermore, mental QOL was best predicted by assessing positive capabilities, such as levels of hope.	Focus on non-western transwomen population	Retrospective cross-sectional study. **Only 4 out of 247 subjects underwent gender-reassignment surgery (GRS)**. The sample size is too small to include GRS status as a predictor in the multivariable regression, and thus **provided no information on the effect of GRS on QOL**.	This study focuses on the transwomen population in one city in China. It is a retrospective cross-sectional study, no before-after comparison. **Only 4 out of 247 subjects underwent gender-reassignment surgery (GRS)**. The sample size is too small to include GRS status as a predictor in the multivariable regression, and thus **provided no information on the effect of GRS on QOL**.
Özkan et al., 2018 DOI: 10.1080/2000656X.2018.1444616	Retrospective	A retrospective study of 43 patients.	Describe the surgical technique, long-term results and sexual outcomes of patients who underwent **vaginal reconstruction with the modified rectosigmoid**	Vaginal reconstruction with denervated rectosigmoid held in an ischemic state appears to be a reasonable option among several available reconstruction techniques.		Retrospective cohort	This is a descriptive study, not a comparative effectiveness study, not providing any information on the effect of treatments
Poteat et al., 2016 Global epidemiology of HIV infection and related syndemics affecting transgender people DOI:	Literature review (but on a different topic)	The citation appears to be incorrect: it is a literature review the HIV epidemic among transgender population, not related to QoL and psychological functions.					**This citation might be incorrect**: it is literature review of the HIV epidemic among transgender population, doesn't seem to related to QoL and psychological functions.

Nobili et al., 2018 DOI: 10.1007/s11154-018-9459-y	Literature review, and formal statistical meta-analysis	Random effects meta-analysis of 29 studies up to July 2017.	A review of quality of life of treatment-seeking transgender adults	Transgender people display poor QoL, independent of the domain investigated. Meta-analysis in a subgroup of studies looking at QoL in participants who were exclusively post-CHT(**Cross-sex Hormonal Treatment**) found no difference in mental health QoL between groups. Insufficient data for a pre-treatment subgroup.			This is a review and meta analysis. It combines several types of treatments (e.g. surgery and hormone therapy). It noted "**the majority of the studies were cross-sectional, lacked controls, and displayed moderate risk of bias…Better quality studies that include clearly defined transgender populations, divided by stage of gender affirming treatment and with appropriate matched control groups are needed to draw firmer conclusions.**"
White Hughto & Reisner, 2016. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. Transgender Health, 1(1), 21–31. https://doi.org/10.1089/trgh.2015.0008.	Literature review (of three prospective studies)	Review of three uncontrolled prospective cohort studies, enrolling 247 transgender adults. The studies measured exposure to hormone therapy and subsequent changes in mental health (e.g., depression, anxiety) and quality of life outcomes at follow-up.	A review of prospective cohort studies to study the effects of **hormone therapy** on psychological functioning and quality of life in transgender individuals	Two studies showed a significant improvement in psychological functioning at 3–6 months and 12 months compared with baseline after initiating hormone therapy. The third study showed improvements in quality of life outcomes 12 months after initiating hormone therapy for FTM and MTF participants; however, only MTF participants showed a statistically significant increase in general quality of life after initiating hormone therapy.	Focus on **prospective cohort** studies		This is a literature review on the effects of **hormone therapy**; no information on the effects of SRS. The paper acknowledge several weaknesses that are common in the transgender literature: "Hormone therapy interventions to improve the mental health and quality of life in transgender people with gender dysphoria **have not been evaluated in controlled trials**. **Low quality evidence** suggests that hormone therapy may lead to improvements in psychological functioning. **Prospective controlled trials are needed** to investigate the effects of hormone therapy on the mental health of transgender people."
T'Sjoen et al., 2019: Endocrinology of transgender medicine. Endocrine Reviews, 40(1), 97–117.https://doi.org/10.1210/er.2018-00011.	Literature review, no meta-analysis	Comphrehensive review of hormal treatments of transgender people	To review recent data on **hormonal treatment** of transgender population and its effect on physical, psychological, and mental health.	A summary of the procedure and outcomes of hormonal treatments of transgender people.		Qualitative summaries, no formal meta-analysis.	This is a literature review. It cited three papers in suporting the statement in abstract of "Mental health problems such as depression and anxiety have been found to reduce considerably following hormonal treatment," but provided no data. The paper acknowledges "current available research is based mostly on cross-sectional studies…long-term follow-up studies and studies involving large groups of people are needed to evaluate whether these improvements remain."

Eftekhar Ardebili et al. (2020) DOI: 10.1186/s12955-020-01510-0	Literature review, simple meta-analysis	**Meta analysis (weighted mean)** of 8 studies, 1099 patients	To conduct a systematic review and meta-analysis about the quality of life (QoL) of individuals during the post **transsexual surgery** period.	The results of this systematic review may support the approaches to transsexuality that facilitates sex reassignment. In this review, the means of quality of life after surgery were not compared to the means of quality of life before surgery or even before hormonal therapy which was due to inadequate number of primary studies		Retrospective design: the means of quality of life after surgery were not compared to the means of quality of life before surgery or even before hormonal therapy which was due to inadequate number of primary studies. **So it provides no information regarding the effect of surgery on the QOL.**	**Meta-analysis**. No pre-surgery information, and thus **provides no information regarding the effect of surgery on the QOL.**
Baker et al. 2021. Hormone therapy, mental health, and quality of life among transgender people: A systematic review. Journal of the Endocrine Society, 5(4),https://doi.org/10.1210/jendso/bvab011.	literature review	Literature review of 20 studies reported in 22 papers that evaluate quality of life (QOL), depression, anxiety, and death by suicide in the context of gender-affirming hormone therapy among transgender people of any age. Fifteen were trials or prospective cohorts, one was a retrospective cohort, and 4 were cross-sectional.	To provide a systematical review the effect of gender-affirming **hormone therapy** on psychological outcomes among transgender people.	Hormone therapy was associated with increased QOL, decreased depression, and decreased anxiety. Associations were similar across gender identity and age. Could not draw any conclusions about death by suicide.			This paper is a literature review focusing on effects of hormonal therapy on QoL and mental health; **it doesn't provide any information on the effects of SRS** (surgery). The authors acknowleges that "this conclusion is limited by high risk of bias in study designs, small sample sizes, and confounding with other interventions."
Javier et al., 2022 DOI: 10.1080/26895269.2022.2038334	Literature review, no meta-analysis	A qualitative literature review of seventy-nine low quality (e.g., small sample sizes, lack of control/comparison groups)	A systematic literature review into the longer-term (i.e., ≥ 1 year) surgical satisfaction and quality of outcomes following various forms of **gender-affirming surgery** in transgender populations.	Seventy-nine low quality studies suggest that most transgender patients are satisfied with surgical outcomes when assessed at least one-year post-surgery. Low quality research also indicates that transgender women and men typically report positive psychological and sexual wellbeing post-surgery, and similar wellbeing outcomes as those who have not had surgery.		**Qualitative summaries, no formal meta-analysis**, **all the studies are low quality** (e.g., small sample sizes, lack of control/comparison groups)	**Conflicting findings but the conclusion gears towards the positive summaries**. Not all longer-term outcomes following GAS reported in this review were positive. A small minority of transgender men and women among the studies reviewed reported low levels of surgical satisfaction (e.g., Ainsworth and Spiegel, 2010; Bouman et al., 2016; Fakin et al., 2019; Leriche et al., 2008). A small minority of transgender women and men also reported regretting having surgery (e.g., Amend et al., 2013; Antoszewski et al., 2012; Garcia et al., 2014; Neuville et al., 2021).

WPATH Assertion # 2

Gender-affirming interventions are based on decades of clinical experience and research; therefore, they are not considered experimental, cosmetic, or for the mere convenience of a patient. They are safe and **effective at reducing gender incongruence and gender dysphoria** (e.g., Aires et al., 2020; Aldridge et al., 2020; Al-Tamimi et al., 2019; Balakrishnan et al., 2020; Baker et al., 2021; Bertrand et al., 2017; Buncamper et al., 2016; Claes et al., 2018; Eftekhar Ardebili, 2020; Esmonde et al., 2019; Javier et al., 2022; Lindqvist et al., 2017; Lo Russo et al., 2017; Marinkovic & Newfield, 2017; Mullins et al., 2021; Nobili et al., 2018; Olson-Kennedy, Rosenthal et al., 2018; Özkan et al., 2018; Poudrier et al., 2019; T'Sjoen et al., 2019; van de Grift, Elaut et al., 2018; White Hughto & Reisner, Poteat et al., 2016; Wierckx, van Caenegem et al., 2014; Wolter et al., 2015; Wolter et al., 2018). Page S18

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes

Aldridge et al., 2020 https://doi.org/10.1111/andr.12884	Prospective, before-after	Participants (n = 178) completed a socio-demographic questionnaire, the Hospital Anxiety and Depression Scale (HADS), the Multidimensional Scale of Perceived Social Support (MSPSS) and the Autism Spectrum Quotient—Short Version (AQ-Short) at pre-assessment (T0) and at 18 months after initiation of GAHT (T1).	To investigate the effect of 18-month GAHT on depression and anxiety symptomatology and the predictors on mental health outcomes in a large population of transgender people, using a longitudinal study	From T0 to T1, symptomatology was significantly decreased for depression (P < .001) and non-significantly reduced for anxiety (P = .37). Scores on the MSPSS predicted reduction in depression, while scores on the AQ-Short predicted reduction in anxiety.	Prospective, before-after comparison		The results are mixed: depression was significantly reduced, but anxiety is not. Even though this is prospective before-after comparison, it is still subject to unmeasured confounding.
Lo Russo et al. 2017	Prospective, but no before-after outcomes	25 FTM transgender patients underwent surgical procedures to create a masculine chest-wall contour. In our study, we just considered 16 patients who have undergone chest surgery with the double incision method. Outcomes are on surgical complications, scar, etc. Outcome on satisfication is aesthetically pleasing and surgeon evaluations.	To propose a new technical approach (chest-wall contouring surgery) and describe indications for FTM transgender patients' surgery.	The patients' survey revealed a high satisfaction rate with the aesthetic result.		No before-after comparison. No outcomes on gender incongruence and gender dysphoria.	The paper focuses on describing a new surgical technique. There is no before-after comparison. The satisfaction outcome is limited to aesthetic results. There is no outcome on gender dysphoria.
Olson-Kennedy, Rosenthal et al., 2018. Health considerations for gender non-conforming children and transgender adolescents. Guidelines for the primary care of	Guidelines						This is part of the "Guidelines for the Primary Care of Transgender and Gender nonbinary people." It is not an individual scientific study.
Marinkovic & Newfield, 2017 https://doi.org/10.1080/15532739.2017.1349706.	Retrospective	Data on 14 patients who underwent chest reconstruction surgery. Outcomes include self-reported satisfication on aesthetics of the surgical outcome and complication	To present data about chest reconstructive surgeries in transgender youth from a Pediatric Gender Management (GeM) clinic.				The paper presents data about chest reconstructive surgeries in transgender youth. The outcomes do not include measures of gender dysphoria, only self-reported satisfaction of the surgical outcome. There is no before-after comparison.

Poudrier et al., 2019 https://doi.org/10.1097/PRS.0000000000005113	Retrospective	An anonymous online survey was distributed to 81 of the senior author's former top-surgery patients. The survey response rate was 72 percent (58 respondents). Responses were analyzed to investigate quality of life, sexual confidence, mental health, satisfaction with top surgery, and patient attitudes toward top surgery's role in gender affirmation.	To assess Quality of Life and Patient-Reported Satisfaction with Masculinizing Top Surgery.	Top surgery had major positive effects on all mental health and quality-of-life metrics.		1. retrospective; 2. nonresponse bias; 3. self-developed survey, self-reported outcomes, not standardized instruments	The paper obtains data from a retrospective cross-sectional survey. The questions are directly on the self-reported improvement of QoL, Sexual Confidence, and Mental Health, but this is not actual measurements of the same outcome before and after the SRS (as would be the case in a prospective before-after design). So it is not an objective before-after comparison, and is subject to recall bias. The response rate is 72% and thus is subject to nonresponse bias. Finally, the outcomes are all self-reported using a self developed survey, and there is no direct measurements on geneder dysphoria.
Wolter et al., 2018 https://doi.org/10.1016/j.bjps.2017.09.003.	Retrospective	Compare an earlier cohort and a later cohort with new preventive measures. The outcomes are complication rate, patient satisfaction and secondary revision rate	To describe and assess new preventive measures in subcutaneous mastectomy in female-to-male transsexuals	By implementation of peri- and postoperative preventive measures and additional application of an algorithmic care path we could achieve a significant reduction of complications, particularly of the hematoma evacuation rate			This paper describes and assesses new preventive measures in subcutaneous mastectomy in female-to-male transsexuals. It does not provide any information on geneder dysphoria.
Wolter et al., 2015 https://doi.org/10.1016/j.bjps.2014.10.016.	Retrospective	The records of 173 patients (346 mastectomies) from were retrospectively reviewed. The authors conducted four different surgical techniques depending on breast volume, grade of ptosis and skin elasticity. The outcome parameters such as complication rate, patient satisfaction with the aesthetic result, nipple sensitivity and surgical correction rate were obtained and related to the employed technique.	To introduce an algorithm to facilitate choosing the appropriate mastectomy technique depending on morphological aspects.	Introduced an algorithm for choosing mastectomy technique			This paper focuses on developing an algorithm to facilitate choosing the appropriate mastectomy technique depending on morphological aspects. There is no information on gender dysphoria.
Bertrand 2017 DOI: 10.1016/j.anplas.2017.05.005	Retrospective cross-sectional	22 patients contacted, 16 respondents, self reported based on questionaire	Evaluate patient satisfaction following **bilateral mastectomy** for female-to-male gender	The mean aesthetic score was 332/378, corresponded to "very satisfied" in the questionnaire. The psychological score was 54.5/60.		1. No comparison groups, only cross sectional summary of a single group who responded, nonresponse bias 2. No information on pre-surgery measurements. 3. Very small sample size	The paper is in French; I only reviewed the English abstract. The study has no comparison groups, only cross-sectional summary of subjects who responded, subject to nonresponse bias. There is no information on pre-surgery measurements and thus no before-after comparison. The sample size is small.

Claes et al., 2018 DOI: 10.1016/j.cps.2018.03.010	No study sample, no design	Description of chest surgery	Describe the surgerical techniques of c**hest surgery** for Transgender and Gender		Description of the clinical aspect of Chest Surgery, **no information on effects on mental and physiology outcomes**. No specific study sample
Esmonde et al., 2019 DOI: 10.1093/asj/sjy166	Retrospective cross-sectional	A retrospective review of 458 patients who underwent chest-affirming procedures in a single institution based on a questionnaire on quality of life and body image outcomes	Describe a single-institution experience of chest-affirming procedures performed in nonbinary patients, including patient characteristics, surgical	Retrospective, no before-after comparison	This paper focuses on describing the surgical techniques of chest affirming surgery. There is no before-after comparisons, and thus provides no data supporting the claim in the abstract that "patients reported improved quality of life."
Aires et al., 2020 DOI: 10.1080/26895269.2020.1848690	Duplicate				Reviewed in Assertion 1. Specific to this Assertion, there is no outcome measures on gender dysphoria or general mental or physical well-being, and thus **provides no direct evidence supporting the Assertion** of "effective in reducing gender incongruence and gender dysphoria"
Al-Tamimi et al., 2019 DOI: 10.1016/j.jsxm.2019.07.027	Duplicate				Reviewed in Assertion 1. Specific to this Assertion, this is a descriptive study, describing the surgical techniques, and reporting on the clinical outcomes; not a comparative study on the effect of the SRS. Specific to this Assertion, there is no outcome measures on gender dysphoria or general mental or physical well-being, and thus **provides no direct evidence** supporting the Assertion of "effective in reducing gender incongruence and gender dysphoria"
Balakrishnan et al., 2020 DOI: 10.1080/26895269.2022.2100644	Duplicate				Reviewed in Assertion 1. Specific to this Assertion, this is a descriptive study of the outcomes of a specific type of surgery; it is not a comparative effectiveness analysis, and there is no outcome measures directly measureing gender dysphoria. Thus it **provides no evidence supporting the Assertion** of "effective in reducing gender incongruence and gender dysphoria"
Buncamper et al., 2016 DOI: 10.1080/26895269.2022.2100644	Duplicate				Reviewed in Assertion 1. Specific to this Assertion, this is a description of the surgical outcomes (e.g. complications) of penile inversion vaginoplasty. It is not a comparative effectiveness analysis, and there is no outcome measures measureing gender dysphoria. Thus it **provides no evidence** supporting the Assertion of "effective in reducing gender incongruence and gender dysphoria"

Eftekhar Ardebili DOI: 10.1186/s12955-020-01510-0	Duplicate	Reviewed in Assertion 1. Specific to this Assertion, this study focuses on QoL of SRS. However, because non of the studies reviewed contain pre-surgery information, the analysis provides no information regarding the compariative effectiveness of surgery on the QOL. Also, the studies reviewed do not contain direct measures on gender dysphoria, and thus provides no evidence supporting the Assertion of "effective in reducing gender incongruence and gender dysphoria"
Javier et al., 2022 DOI: 10.1080/26895269.2022.2038334	Duplicate	Reviewed in Assertion 1. It is a qualitative summary of 97 low quality studies and found conflicting findings. Indeed, not all longer-term outcomes following GAS reported in this review were positive.
Lindqvist et al., 2017 DOI: 10.1007/s00238-016-1252-0	Duplicate	Reviewed in Assertion 1. It is one of the few prospective studies with before-after SRS data of the same patients. Specific to this Assertion, this paper in fact shows that 3 and 5 years post-operation the QoL is worse than pre-operation. This paper focuses on general QoL outcome and provides no data on direct measure of gender dysphoria. Thus it provides no direct evidence supporting the Assertion of "effective in reducing gender incongruence and gender dysphoria"
Nobili et al., 2018 DOI: 10.1007/s11154-018-9459-y	Duplicate	Reviewed in Assertion 1. It is a meta-analysis of 29 studies, focusing on the outcome of QoL. It does not study the safety of the treatments, and does not provide any information on direct measure of gender dysphoria, no before-after comparison. Thus it **provides no direct evidence supporting the Assertion** of "effective in reducing gender incongruence and gender dysphoria"

van de Grift, Elaut et al., 2018 DOI: 10.1080/0092623X.2017.1326190	Duplicate							Reviewed in Assertion 1. Specific to this assertion, the study has measures on various post-SRS outcomes (e.g. complications, satisfaction, QoL, happiness, psychological symptoms). However, except for psychological symptoms, it didn't collect before-after data on any other outcomes, and thus could not conduct a direct before-after comparison of these outcomes of the same patients. Therefore, it **provides no direct evidence supporting the Assertion** of "effective in reducing gender incongruence and gender dysphoria"
White Hughto & Reisner, Poteat et al., 2016 DOI: 10.1097/QAI.0000000000001087	Duplicate							Reviewed in Assertion 1. Specific to this assertion, it focuses on hormone treatments and provides no information on the effects of SRS. It also calls for "Prospective controlled trials"
Baker et al., 2021	duplicate							Reviewed in Assertion 1. Specific to this assertion, it doesn't provide any information on the effects of GAT
Mullins et al., 2021	duplicate							Reviewed in Assertion 1. Specific to this assertion, this paper does not provide any information on the effects of GAT on gender dysphoria, quality of life and well-being.
T'Sjoen et al., 2019	duplicate							Reviewed in Assertion 1. Specific to this assertion, this literature review does not provide any data in supporting the statement in abstract of "Mental health problems such as depression and anxiety have been found to reduce considerably following hormonal treatment." The outcomes do not directly measure gender dysphoria.
Özkan et al., 2018	Duplicate							Reviewed in Assertion 1. Specific to this assertion, this is a descriptive study, not a comparative effectiveness study, not providing any information on the effect of GAT
Wierckx, van Caenegem et al., 2014	Duplicate							Reviewed in Assertion 1. Specific to this assertion, this paper does not provide information on the effects on gender dysphoria.

WPATH Assertion # 3							
However, the recommendations put forth here apply to all institutions that house TGD individuals, both carceral and noncarceral (Porter et al., 2016). Page S104							
Name	**Study-Design**	**Methodology**	**Aim/Objective**	**Conclusions**	**Strengths**	**Limitations**	**Notes**

Porter, 2016 DOI: 10.1080/07317115.2016.1203383	Literature review		To provide a context from which clinicians from all disciplines, not only psychology, can apply the American Psychological Association (APA) 2015 Guidelines to provide competent and affirming services to older (transgender and gender nonconforming) TGNC	Interpret the APA Guidelines using a gerontological lens to elucidate specific issues faced by the TGNC older adult along with the practice and policy implications for this population.			This paper focuses on older TGNC adults. Its applicability to general TGNC population is unclear

WPATH Assertion # 4

People should have access to these medically necessary treatments irrespective of their housing situation within an institution (Brown, 2009). Page S104

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Brown, 2009 DOI: 10.1080/15532730903008073	Opinion						This is an opinion piece (specifically, recommended revisions to WPAH Standards of Care Section on Medical Care for Incarcerated Persons with Gender Identity Disorder). It is not a scientific study; there is no study design or methodology or data.

WPATH Assertion # 5

TGD residents in carceral facilities report the lack of access to medically necessary transgender-specific health care (see Chapter 2—Global Applicability, Statement 2.1), which is ranked as their number one concern while incarcerated (Brown, 2014; Emmer et al., 2011). Page S104

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Brown, 2014 DOI: 10.177/1078345814541533	Retrospective cross-sectional	Qualitative analysis of 129 transgender inmates' correspondence	To identify transgender inmates' concerns	Claims of inadequate care for transgendered patients that have sufficient merit to be fully litigated in U.S. courts appear likely to produce verdicts in favor of plaintiff inmates			The paper **provides evidence for the Assertion**, on page 336, it states "transgender health care issues accounting for the largest number (55%) (of concerns) by a wide margin"
Emmer et al., 2011: Emmer, P., Lowe, A., & Marshall, R. B. (2011). This is a prison, glitter is not allowed : Experiences of trans and gender variant people in	Retrospective cross-sectional	Self-developed survey of TG immates, respondent rate is 68 out of over 100.	To provide a report on the experiences of trans and gender variant people in Pennsylvania's Prison System				The report found "Requests for and delivery of gender-related health services were often met with ignorance and ntimidation."

WPATH Assertion # 6

Controlled studies show clinically significant health and mental health disparities for justice-involved transgender people compared to matched groups of transgender people who have not been incarcerated or jailed. (Brown and Jones, 2015). Page S104

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes

Brown and Jones, 2015 DOI: 10.1089/lgbt.2015.0052	Retrospective cross-sectional, matched case-control	Studied a large cohort of transgender TG veterans who received care in Veterans Health Administration (VHA) facilities during 2007–2013 (n = 4,793) and a 3:1 matched control group of veterans without known TG identification (n = 13,625). Three hundred twenty six (326: 138 TG, 188 TG) had received VHA services in programs designed to address the needs of justice involved veterans. Linked patients in each of the three groups to their medical and administrative data.	To investigate health disparities in transgender veterans involved with the criminal justice system	TG veterans experience a number of health risks compared to non-TG veterans, including an increased likelihood of justice involvement.	matched case-control study	focus on U.S. veterans, generalizability to general population is unclear.	The study compares health disparities between **TG and non-TG** veterans, and found TG veterans have an increased likelihood of justice involvement. **This comparison is entirely different from the comparison stated in the Assertion**, namely, **justice-involved TG people vs. non justice-involved TG people** (i.e. comparison between different TG populations). Therefore, the study is not relevant to the claim in this Assertion.

WPATH Assertion # 7

Too often the agencies, structures, and personnel that provide care are lacking in knowledge, training, and capacity to care for gender diverse people (Clark et al., 2017) Page S104

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Clark et al., 2017 DOI: 10.1016/j.socscimed.2017.09.052.	Cross-sectional interview	Interview of 20 correctional healthcare providers (e.g., physicians, social workers, psychologists, mental health counselors) from New England	To examine correctional healthcare providers' knowledge of, attitudes toward, and experiences providing care to transgender inmates.	Transgender inmates do not consistently receive adequate or gender-affirming care while incarcerated.		Small sample size in one region.	The paper provided facts supporting the Assertion. The limitation is that the sample size is small (20) and is restricted to one region (New England).

WPATH Assertion # 8

We recommend the application of the Standards of Care (SOC) to people living in institutions as basic principles of health care and ethics (Beauchamp & Childress, 2019; Pope & Vasquez, 2016). S105

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Beauchamp & Childress, 2019 Am J Bioethics	Opinion						This is an opinion piece (guest editorial of a journal). It is not a scientific study; there is no study design or methodology or data.
Pope & Vasquez, 2016: Ethics in psychotherapy and counseling: A practical guide	Book						This is a 496 page long book. It made the call as stated in the Assertation, which is an opinion (recommendation of care). There is no study design or methodology or data.

WPATH Assertion # 9

TGD people with gender dysphoria should have an appropriate treatment plan to provide medically necessary surgical treatments that contain similar elements provided to persons who reside outside institutions (Adams v. Federal Bureau of Prisons, No. 09-10272 [D. MO June 7, 2010]; Brown 2009; Edmo v. Idaho Department of Corrections, 2020). Page S106

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes

Brown, 2009 DOI: 10.1080/1553273090300	Duplicate						Reviewed in Assertion 4. It is an opinion piece, not a scientific study; there is no study design or methodology or data.

WPATH Assertion # 10

Gender-affirming vaginoplasty is one of the most frequently reported gender-affirming surgical interventions; 8 prospective (Buncamper et al., 2017; Cardoso da Silva et al., 2016; Kanhai, 2016; Manero Vazquez et al., 2018; Papadopulos, Zavlin et al., 2017; Tavakkoli Tabassi et al., 2015; Wei et al., 2018; Zavlin et al., 2018), 15 retrospective cohort (Bouman, van der Sluis et al., 2016; Buncamper et al., 2015; Hess et al., 2016; Jiang et al., 2018; LeBreton et al., 2017; Manrique et al., 2018; Massie et al., 2018; Morrison et al., 2015; Papadopulos, Lelle et al., 2017; Raigosa et al., 2015; Salgado et al., 2018; Seyed-Forootan et al., 2018; Sigurjonsson et al., 2017; Simonsen et al., 2016; Thalaivirithan et al., 2018), and 3 cross-sectional cohort studies have recently been reported (Castellano et al., 2015; Owen-Smith et al., 2018; van de Grift, Elaut et al., 2018). Page S128

Although different assessment measurements were used, the results from all studies **consistently reported both a high level of patient satisfaction** (78–100%) as well as satisfaction with sexual function (75–100%). This was especially evident when using more recent surgical techniques. Gender-affirming vaginoplasty was also associated with a low rate of complications and a low incidence of regret (0–8%). S128-129

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Buncamper et al., 2017 DOI: 10.1097/PRS.0000000000003108	Prospective	Prospective study of 100 patients	provide data to enable an evidence-based discussion on the controversy of whether or not to use additional full-thickness skin graft in penile inversion vaginoplasty.	The authors can confirm neither of the suggested arguments, for or against full-thickness skin graft use, in penile inversion vaginoplasty.			This paper focuses on the debate of a specific surgical detail, and analyzed the association between the surgical technique and the patient-reported aesthetic outcome, overall satisfaction with the neovagina, sexual function, and genital self-image, and found no assocations. There is no discussion on QoL, necessity, statisfaction of any particular technique.
Kanhai, 2016 DOI: 10.1007/s00266-016-0620-2	Prospective, but no before-after outcome	Prospective study of 50 patients	Describe Sensate pedicled-spot plasty: a new technique to create a sexual arousal organ in the anterior wall of the	Sensate pedicled-spot plasty is a safe innovative technique which lead to adequate sexual functionality in all patients.			Description of a new technique in SRS. Nothing about pre- and post- SRS comparison. No outcome measures other than sexuality functionality
Manero Vazquez et al., 2018 DOI: 10.1097/SAP.0000000000001532	Prospective, but no before-after outcome	97 Ninety-seven patients in one medical center. All clitoroplasties and vulvoplasties were completed in the same surgical stage as the vaginoplasty	describe a new clitoroplasty and vulvoplasty technique in male-to-female sex reassignment surgery and its outcome.				Description of the authors' clitoroplasty and vulvoplasty technique. Outcomes are all clinical outcomes, n**o information on other outcome measures such as QoL.**

Papadopulos, Zavlin et al., 2017 DOI: 10.1097/PRS.0000000000003529	Prospective, before-after	Prospective study of **39** patients, QoL based on self-developed questions as well as standarized questionaries during both pre- and post- operative (6 months) periods	**Evaluate QoL among MTF SRS using combined technique**	Statistically significant improvements were found in the paired T test (pre- and post-SRS) regarding Questions on Life Satisfaction, Modules (German version), especially for the items "partnership," "ability to relax," "energy," "freedom from "anxiety," "hair," "breast," and penis/vagina" ($p < 0.01$). Furthermore, the patients appeared more emotionally stable ($p = 0.03$), showed higher self-esteem ($p = 0.01$), and showed much lower depression/anxiety ($p < 0.01$).	**Prospective**: information of the same patient pre- and post- (6 months) SRS.	Small sample size, short term follow-up	One of the few studies focusing on QoL and with different comparison pre-and post- SRS. A **multiple testing** issue: too many hypothesis testing (41) on too few patients (39). Multiple testing is a serious statistical program in hypothesis testing: the more tests you conduct on the same set of data, the more likely you will have false positives. For example, if you use p=0.05 as a threshold hold, among 100 tests, you will expect to have at least 5 false positives just by random chance. This is not specific to a single type of SRS, but a combined SRS technique.
Tavakkoli Tabassi et al., 2015 DOI: 10.1007/s00238-014-1038-1	Prospective, but no before-after outcome	112 previously circumcised MTF subjects underwent a modification of penile inversion vaginoplasty	Description of a surgical technique of male-to-female (MTF) transsexual vaginoplasty when subjects have short penile skin flap because of circumcision				Description a special type of vaginoplasty. **No comparison groups. No information on other outcome measures, e.g. qualify of life. Didn't address the necessity of the surgery**. Only described the satisfaction rate of the patients of a special type of technique.
Wei et al., 2018 DOI: 10.1007/s00266-017-0977-x	Prospective, but no before-after outcome	Nine male-to-female transsexual patients received our new method of vaginoplasty from July 2010–October 2015. mean clinical follow-up period of 25.3 months and phone interview follow-up of 50.3 months	described the details of the surgical procedure and evaluated the long-term anatomical and functional outcomes.	All the patients were sexually active and reported sexual satisfaction		Small sample size, no comparison between pre- and post- surgery	Description a new technique of vaginoplasty. **Focuses on long-term anatomical and functional outcomes. No comparison groups. No information on other outcome measures, e.g. qualify of life. Didn't address the necessity of the surgery**.
Zavlin et al., 2018 Aesthetic Plastic Surgery, 42(1), 178–187. https://doi.org/10.1007/s00266-017-1003-z.	Prospective, but no before-after outcome	A prospective study of 40 MTF patients undergoing SRS using **Combined Vaginoplasty**. The outcome are measured using self-developed (by the authors) indication-specific questionnaires to evaluate the aesthetic, functional, and sexual outcomes of SRS 1 day before stage 0 and 6 months after stage 1.	Assess Satisfaction of MTF patients undergoing SRS using Combined Vaginoplasty.			1. no comparison groups; 2. standardized and validated SRS-specific questionnaires are lacking; 3. short follow up. 4. small sample size	No information on qualify of life. **No comparison groups; didn't address the necessity of the surgery. No standardized and validated SRS-specific questionnaires**

Bouman, van der Sluis et al., 2016 DOI: 10.1016/j.jsxm.2016.06.009	Retrospective	A retrospective survey of 31 transwomen	Assess patient-reported functional and esthetic outcomes, quality of life, satisfaction, and sexual well-being after **primary total laparoscopic intestinal vaginoplasty** in transgender women.	Low Female Sexual Function Index scores (FSFI); satisfactory functional and esthetic results of the neovagina and a good quality of life		restrospective study, no pre- and post- SRS comparison, small sample size	No before and after comparison, small sample size, self-reported outcome. low score on sexual functionality, which is consistent with the founding in Buncamper et al. 2015.
Buncamper et al., 2015 DOI: 10.1111/jsm.12914	Retrospective	Retrospective survey study on 49 transgender women. Primary outcomes were self-reported functional and aesthetic evaluation. Secondary outcomes were the aesthetic evaluation of the vaginoplasty by an independent panel.	Assess if **penile skin inversion neovaginoplasty** performed in transgender women achievers satisfactory functional and aesthetic outcomes, as well as the physical and sexual well-being and	56% patients are sexually dysfunctional according to the Female Sexual Function Index (FSFI) score. Majority of the patients are satisfied with both the functional and aesthetic results of neovaginoplasty using penile skin inversion			No comparison groups. No information on other outcome measures, e.g. qualify of life. Didn't address the necessity of the surgery. Indeed recorded that **56% patients are sexually dysfunctional following the surgery.**
Hess et al., 2016 DOI: 10.1159/000443281	Retrospective	Retrospective review of 96 MtF transgender patients in a single Brazil medical center	Assesses sensitivity of the neoclitoris following a new preparation in MTF SRS.				Description of the surgical procedure (Modified Preparation of the Neurovascular Bundle in MTF patients). **No comparison groups. No information on other outcome measures, e.g. qualify of life. Didn't address the necessity of the surgery.**
Jiang et al., 2018 DOI: 10.1016/j.jsxm.2018.03.085	Retrospective	486 patients were seen in consultation for trans-feminine gender-affirming genital surgery: 396 requested vaginoplasty and 39 patients requested vulvoplasty. 30 Patients either completed or are scheduled for vulvoplasty	describe the factors influencing patient choice or surgeon recommendation of **vulvoplasty** and to assess the patient's satisfaction with this choice	Vulvoplasty patients were older and had higher body mass index than those seeking vaginoplasty. The majority (63%) of the patients seeking vulvoplasty chose this surgery despite no contra-indications to vaginoplasty. The remaining patients had risk factors leading the surgeon to recommend vulvoplasty. Of those who completed surgery, 93% were satisfied with the surgery and their decision for vulvoplasty		retrospective, non-validated questions, short-term follow-up, and selection bias in how we offer vulvoplasty	Description of the factors influencing patient choice or surgeon recommendation of vulvoplasty and to assess the patient's satisfaction with this choice. **No comparison groups. No information on other outcome measures, e.g. qualify of life. Didn't address the necessity of the surgery.**

LeBreton et al., 2017 DOI: 10.1016/j.jsxm.2016.12.005	Prospective, but no before-after outcome	Prospective study on 28 transgender women at least 18 years old operated on at least 3 months before the study by a single surgeon. Outcomes are clinical outcomes such as medical complications, and self reported questionnaires on general and sexual satisfaction, sexual function, depression, and psychological well-being.	investigate genital sensory detection thresholds in male-to-female transgender women postoperatively (**Vaginoplasty**) and their relation to psychological well-being and variables of satisfaction	Gender-affirming surgery yields good results for satisfaction with appearance and function. Genital sensitivity showed the best results with pressure and vibration		No pre- and post- SRS comparison of the same patient	**Despite being a prospective study, only provided one item to assess sexual satisfaction before and after GAS**. All other outcomes are measured post-SRS. Small sample size; self-reported outcome on statisfication. Focus on genital sensations
Manrique et al., 2018 DOI: 10.1097/PRS.0000000000004122	Retrospective	retrospective chart review of **15** transgender women who underwent gender-confirmation surgery using the **pedicle transverse colon flap**.	present the clinical outcomes and sexual function evaluation when using the pedicle transverse colon flap for gender-confirmation			restrospective, no before-after comparison. small sample size (15)	This paper focus on the clinical outcomes and sexual function evaluation. No discussion on satisfaction or quality of life. No before- and after-comparison.
Massie et al., 2018. Predictors of Patient Satisfaction and Postoperative Complications in Penile Inversion Vaginoplasty DOI: 10.1097/PRS.0000000000004427	Retrospective	retrospective chart review of 117 patients from a single surgeon's experience with **penile inversion vaginoplasty** (performed between July of 2014 to June of 2016). Report on both postoperative complications and patient-reported outcomes.	report both postoperative complications and patient-reported outcomes from the largest cohort in the United States to date to undergo penile inversion vaginoplasty	Most common complications were granulation tissue (26 percent), intravaginal scarring (20 percent), and prolonged pain (20 percent). 94 percent patients reported "feeling positively about their genitals" and 94 percent reported "would do this operation again"). **Seventy-one percent of patients reported resolution of their gender dysphoria**.		restrospective study, no pre- and post- SRS comparison	It is restrospective study. **No pre- and post- SRS comparison.** Patient statisfication is self-reported.
Morrison et al., 2015 DOI: 10.1097/PRS.0000000000001459	Retrospective	A retrospective review of 83 MTF patients who had undergone **rectosigmoid neocolporrhaphy** by a single doctor over 22 years	provide an objective investigation into the safety and efficacy of rectosigmoid neocolporrhaphy for vaginoplasty in male-to-female transsexual patients.	Rectosigmoid vaginoplasties are safe and effective, and have good long-term results in primary and secondary vaginoplasties. Forty-eight patients (58 percent) had complications, but the majority (83.3 percent) were minor and consisted mainly of introital stricture or excessive protrusion of the corpus spongiosum. Overall patient satisfaction with appearance and sexual function was high.		restrospective study, no pre- and post- SRS comparison	Restrospective study, no pre- and post- SRS comparison. Focus on a specific type of vaginoplasty.

Papadopulos, Lelle et al., 2017 DOI: 10.1016/j.jsxm.2017.01.022	Retrospective	Retrospective survey of 47 patients after SRS. Outcomes are reported on self-developed indication-specific questionnaire on postoperative satisfaction and a standardized self-assessment questionnaire (FLZ) on satisfaction and QOL	Analyze patient satisfaction and QOL after SRS.	Self-developed indication-specific questionnaire showed that 91% experienced an improvement of QOL. or the FLZ, the sum score for general life satisfaction (P < .001) was significantly lower than the normative data, whereas the sum score of the satisfaction with health module (P = .038) did not reach statistical significance.		restrospective study, no pre- and post- SRS comparison, subjective to recall bias	restrospective study, no pre- and post- SRS comparison
Raigosa et al., 2015 DOI: 10.1111/jsm.12936	Retrospective cross-sectional	A Retrospective Review of Surgical Technique and Complications in 60 Patients	Review of Surgical Technique and Complications in MTF genital reassignment surgery (GRS)	GRS can provide good functional and aesthetic outcomes in patients with male-to-female GD. However, despite a careful planning and meticulous surgical technique, secondary procedures are frequently required to improve the function and appearance of the neovagina.			This paper focuses on reviewing the surgical technique, clinical outcomes and complications of GRS, no information on other outcomes such as QoL, satisfication, or necessity of the GRS
Salgado et al., 2018 DOI: 10.1155/2018/4907208	Retrospective cross-sectional	A retrospective review of **12** patients in a single hospital	To describe the surgical technique and outcomes in primary sigmoid vaginoplasty in transwomen	Sigmoid vaginoplasty is a reliable technique for achieving a satisfactory vaginal depth that is sexually functional.			The study focuses on description of surgical techniques and functional outcomes. No data on patient satisification, no before-after comparison
Seyed-Forootan et al., 2018 DOI: 10.1007/s00266-018-1088-z	Retrospective cross-sectional	A retrospective comparison of 24 male-to-female transsexual patients based on their complications and levels of satisfaction (16 patients received amnion grafts with fibroblasts, and 8 patients received only amnion grafts without any additional cellular lining.)	compare the results of amnion grafts with amnion seeded with autograft fibroblasts in Reconstruction of Neo-vagina in Male-to-Female Reassignment Surgery.	The creation of a neo-vaginal canal and its lining with allograft amnion and seeded autologous fibroblasts is an effective method for imitating a normal vagina.		Retrospective cross-sectional	This paper focuses on comparing two specific surgical techniques in reconstruction of neo-vagina in MTF reassignment surgery. It does not compare with other surgical techniques. The outcomes are most clinical outcomes (e.g. the size of neo-vagina, secretion, sensation). No outcomes are on satisfication or QoL
Sigurjonsson et al., 2017 DOI: 10.1016/j.jsxm.2016.12.003	Retrospective	A retrospective review of **22** patients, with a mean follow-up of 37 months after initial surgery. Main outcome measures are Tactile and vibratory sensitivities of the neoclitoris and questionnaire on satisfaction with orgasm, sexual function, and general satisfaction.	To examine the sensitivity of the neoclitoris and its relation to orgasm and sexual function at least 1 year after GRS.	The vast majority of patients who undergo male-to female GRS experience orgasm and are satisfied with the surgery.	follow up is long term (37 month on average)	Retrospective, no before-after comparison.	This paper focuses on the clinical/functional outcome of SRS, only one summary question about the satisfaction of the SRS, no before-after comparison, no data on general health or QoL.

Simonsen et al., 2016 DOI: 10.3109/08039488.2015.1081405	Retrospective	A retrospective registery study of 104 sex-reassigned individuals in Demark.	(1) To investigate **psychiatric morbidity before and after sex reassignment surgery (SRS)** among Danish individuals who underwent SRS during 1978-2010., (2) To investigate mortality among Danish individuals who underwent SRS during 1978-2010.		Long term natioal sample: the sample comprised 98% (n = 104) of all transexual individuals in Denmark. The outcomes are identified from national registries (Danish Psychiatric Central Research	Despite the over-representation of psychiatric diagnoses both pre- and post-SRS the study found that only a relatively limited number of individuals had received diagnoses both prior to and after SRS. **This suggests that generally SRS may reduce psychological morbidity for some individuals while increasing it for others.**	This study has several strengths: (1) exhaustive national sample over a long period of time; (2) outcome from national registries, not self reported. The results are inconclusive, as the authors concluded "**generally SRS may reduce psychological morbidity for some individuals while increasing it for others**."
Thalaivirithan et al., 2018 DOI: 10.4103/ijps.IJPS_62_18	Retrospective	Retrospective review of 30 transwomen who underwent embryonic equivalents-based sex reassignment surgery (MFEEbSRS) by two independent surgeons	To evaluate the esthetic and functional outcome of **embryonic equivalents-based male-to-female sex reassignment surgery** in transwomen using the institutional scoring	The aesthetic and functional outcome in all the patients was good.			The study focuses on aesthetic and functional outcomes. No data on satistification, QoL, mental well-being. No before-after comparison.
Castellano et al., 2015 DOI: 10.1007/s40618-015-0398-0	Retrospective, case-control	Case-control of 60 pairs of transsexuals and healthy people in Italy. Outcomes measures are testosterone,estradiol, LH and World Health Organization Quality of Life (WHOQOL-100) self-reported questionnaire. Student's t test was applied to compare transsexuals and controls. Multiple regression model was applied to evaluate WHOQOL's chosen items and LH.	To gather information on QoL, quality of sexual life and body image in transpeople at least 2 years after SRS,to compare these results with a control group and to evaluate the relations between the chosen items and hormonal status.	This study highlights general satisfaction after SRS. In particular, transpeople's QoL turns out to be similar to Italian matched controls. LH resulted inversely correlated to body image and sexual life scores.	some confounding control via the case-control design	retrospective, no before-after information of the same subject.	This study is the only matched case-control study in the cited references here. There is no before-after comparison of the same patients, only compared transexual people and healthy people. The controls are matched only a few demographic and SES covariates, can not rule out unmeasured confounding (e.g. mental well-being before SRS, family background, work environment).
Cardoso da Silva et al., 2016 DOI: 10.1016/j.jsxm.2016.03.370	Duplicate (prospective, before-after)						Reviewed in Assertion 1. This is one of the few prospective studies which provide pre- and post-surgery comparison of the same patients. Specific to this assertion, the results on the QoL measures are mixed. In particular, it found "domains I (physical health) and III (level of independence) were significantly worse after SRS." So it is not as clear cut as stated in the Assertion.

Owen-Smith et al., 2018 DOI: 10.1016/j.jsxm.2018.01.017	Duplicate (retrospective)					Reviewed in Assertion 1. Besides the methodological limitations pointed out earlier, a limitation specific to this assertion is that the paper did not directly report the satisfaction rate of patients population and only compared non-treatment vs. treatment groups. Also, the outcome measures (include body-gender congruence, body image satisfaction, depression, and anxiety) do not include measures on physical and sexual functions
van de Grift, Elaut et al., 2018 DOI: 10.1080/0092623X.2017.1326190	Duplicate (prospective, no before-after)					Reviewed in Assertion 1. Specific to this assertion, the paper indeed reports high satisfication rate among SRS patients, but is subject to nonresponse bias because of the low respondent rate; it also has a substantial portion of missing data

WPATH Assertion # 11

Gender-affirming surgical procedures have been shown to relieve symptoms of gender dysphoria and improve mental health (Owen-Smith et al., 2018; van de Grift, Elaut et al., 2017). Page S173

Name	Study-Design/Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Owen-Smith et al., 2018 DOI: 10.1016/j.jsxm.2018.01.017	duplicate					Reviewed in Assertion 1, 10. Specific to this assertion, the study didn't have before-after SRS comparison. It focuses on the comparison between people who received SRS vs not re-received SRS. As commented earlier, such comparison is subject to severe confounding, and does not provide direct evidence for the Assertion of "relieve symptoms of gender dysphoria and improve mental health"
van de Grift, Elaut et al., 2017 DOI: 10.1097%2FPSY.0000000000000465	duplicate					Reviewed in Assertion 1, 2, 10. Specific to this assertion, the study has measures on various post-SRS outcomes (e.g. complications, satisfaction, QoL, happiness, psychological symptoms). However, except for psychological symptoms, it didn't collect before-after data on any other outcomes, and thus could not conduct a direct before-after comparison of these outcomes of the same patients. Therefore, it provides no direct evidence supporting the Assertion of "relieve symptoms of gender dysphoria and improve mental health"

Ettner Assertion # 1

Decades of careful and methodologically sound scientific research have demonstrated that gender-affirming surgeries are safe and effective treatments for severe gender dysphoria and, indeed, **for many people suffering from gender dysphoria, the only effective treatment**. Ettner Report Para. 50.							
Name	**Study-Design**	**Methodology**	**Aim/Objective**	**Conclusions**	**Strengths**	**Limitations**	**Notes**
Pfafflin & Junge (1998) https://www.researchgate.net/publication/299412537_Sex_Reassignment_Thirty_Years_of_International_Follow-up_Studies_after_Sex_Reassignment_Surgery_A_Comprehensive_Review_1961-1991	Literature review	Literature review of thirty years of international follow-up studies of approximately two thousand persons who have undergone sex reassignment surgery in 1961-1991. It includes more than seventy individual studies and eight previously published reviews from many countries and four continents	A comprehensive literature review of thirty years of international follow-up studies of approximately two thousand persons who have undergone sex reassignment surgery.				This is a 300 page book in German, reviewing a large literature that involves 2000 person who underwent SRS between 1961-1991. I reviewed the 6 page translated English abstract. The abstract states " Sex reassignment, properly indicated and performed, has proven to be a valuable tool in the treatment of individuals with transgenderism. It is, however, **not the only powerful change agent in sex reassignment**." So this **directly contradicts** the Assertion of "for many people suffering from gender dysphoria, the only effective treatment"
Smith et al., (2005) DOI:10.1017/s0033291704002776	Prospective	A prospective study of 162 (out of 325 applicants) transexual patients who underwent SRS in the Netherlands. Gender dysphoria was measured with the Utrecht Gender Dysphoria Scale. Follow-up time is 1-4 years	To study the outcomes of sex reassignment, potential differences between subgroups of transsexuals, and predictors of treatment course and outcome.	The results substantiate previous conclusions that sex reassignment is effective. Still, clinicians need to be alert for non-homosexual male-to-females with unfavourable psychological functioning and physical appearance and inconsistent gender dysphoria reports, as these are risk factors for dropping out and poor post-operative results.	prospective		This study uses **before-after SRS comparison of the same patients** to provide evidence for SRS in reducing gender dysphoria as well as improvement in various outcomes on body image and psychological functioning. However, the paper focused solely on the population who underwent SRS, it **provides no evidence** for the Assertion of "for many people suffering from gender dysphoria, the only effective treatment." It is also dated (published in 2005). It is unclear whether the conclusion's applicability today.
Jarolim (2009) - DOI: 10.1111/j.1743-6109.2009.01245.x	Retrospective	A retrospective **3-month follow-up** study of MTF patients' opinions following SRS in 129 patients having a primary procedure. Main outcomes are sexual functions and complications 3 months after surgery. The surgical techniques are described in detail.	To evaluate the results of surgical reassignment of genitalia in male-to-female transsexuals.			Retrospective	The paper focuses on describing the surgical details and outcomes of sexual functions and complications. It did not discuss the effect SRS on gender dysphoria, **nor did it provided any evidence for the Assertion** of "for many people suffering from gender dysphoria, the only effective treatment."

Ettner Assertion # 2						
In 2007, Gijs and Brewayes analyzed 18 studies published between 1990 and 2007, encompassing 807 patients. The researchers concluded: "Summarizing the results from the 18 outcome studies of the last two decades, the conclusion that [gender-affirming surgery] is the most appropriate treatment to alleviate the suffering of extremely gender dysphoric individuals still stands: Ninety-six percent of the persons who underwent [surgery] were satisfied and regret was rare." Ettner Report Para.						
Name	**Study-Design/Methodology**	**Aim/Objective**	**Conclusions**	**Strengths**	**Limitations**	**Notes**

Gijs, L., & Brewaeys, A. (2007).	Literature review	Literature review of 18 studies, involving both MTF and FTM, including many reviewed here (e.g. Smith et al. 2005; Lawrance, 2003, Lobato, et al. 2006; Rehman, et al. 1999; Hepp et al., 2002; De Cuypere, et al. 2005). Qualitative summary	To provide a literature review of SRS in transgender people				Immediately after stating the statement quote in the Assertion (p. 215), the authors acknowledged that "However, **even today this conclusion is based on methodologically less than perfectly designed studies**." Specifically, the paper wrote "Not one of the reviewed outcome studies was a controlled one...In many studies, sound psychometric instruments were not used. **Especially disturbing is that many researchers did not directly measure gender dysphoria as the main outcome variable but instead used derivative measures**, for example, satisfaction with surgery, sexual and interpersonal relationships, occupational and global functioning, or quality of life in general." It also acknowledges a few other methodological shortcomings, like attrition or selection bias of the patient sample.

Ettner Assertion # 3

In 2019, Cornell University published a literature review called What Does the Scholarly Research Say about the Effect of Gender Transition on Transgender Well-Being? The researchers enumerated the following conclusions (omitted here due to length): Ettner Report Para. 56.

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Cornell University What We Know: The Public Policy Research Portal (2019). https://whatweknow.inequality.cornell.edu/topics/lgbt-equality/what-does-the-scholarly-research-say-about-the-well-being-of-transgender-people/	Website						This website collects 51 studies, about half of which are reviewed here. The strengths and weaknesses have been discussed in our review. Overall, the vast majority of the studies in this collection suffers from many methodogical problems, as pointed out also in multiple literature reviews (e.g. Gijs and Brewaeys, 2007; T'Sjoen et al., 2019; White Hughto & Reisner, 2016; Nobilli et al. 2018). Many results are not as clear cut or of direct evidence as stated in the assertion.
Colizzi, M. , Costa, R., & Todarello (2014)	Prospective	A prospective longitudinal study of 118 patients underwent **hormone therapy**	To evaluate the presence of psychiatric diseases/symptoms in transsexual patients and to compare psychiatric distress related to the **hormonal intervention** in a one year follow-up	Hormonal treatment seemed to have a positive effect on transsexual patients' mental health.	prospective study		This study focuses on hormone therapy, not directly related to SRS.

Ettner Assertion # 4							
studies conducted in countries throughout the world likewise conclude that gender-affirming surgery is an extremely effective treatment for gender dysphoria. For example, a 2001 study published in Sweden states: "The vast majority of studies addressing outcome have provided convincing evidence for the benefit of 27 [gender-affirming] surgery in carefully selected cases." Ettner Report Para. 57.							
Name	**Study-Design**	**Methodology**	**Aim/Objective**	**Conclusions**	**Strengths**	**Limitations**	**Notes**
Landen, M. et al. (2001). Done is done – and gone is gone: Sex reassignment surgery is presently the best cure for transsexuals. Lakartidningen, 98(30-31), 3322–26.	Retrospective	Article in Swedish, only reviewed English abstract, which provides no information on study design or methodology.	To summarize the state of the art regarding work-up and treatment of transsexuals.			Small sample size.	This article is in Swedish. I couldn't find the original paper and thus only reviewed the English abstract, which does not contains any information about the study design, methods and outcomes. The paper states indeed stated that the evidence for the benefit of 27 [gender-affirming] surgery is in "**carefully selected cases**." The study is dated (2001) and short; most later literature review (e.g. Gijs, L., & Brewaeys, A. 2007) acknowledges caution about the methodology and conclusion in this field

Ettner Assertion # 5							
Patient satisfaction is an important measure of effective treatment. Achieving functional and normal physical appearance consistent with gender identity alleviates the suffering of gender dysphoria and enables the patient to function in everyday life. Studies have shown that by alleviating the suffering and dysfunction caused by severe gender dysphoria, gender-affirming surgery improves virtually every facet of a patient's life. This includes**satisfaction with interpersonal relationships and improved social functioning**. Ettner Report Para. 58.							
Name	**Study-Design**	**Methodology**	**Aim/Objective**	**Conclusions**	**Strengths**	**Limitations**	**Notes**
Rehman, J., Lazar, S., Benet, A., Schaefer, L, & Melman, A. (1999).	Retrospective	A retrospective review of self reported sex and surgery satisfaction among 28 MTF transsexual patients post SRS. Self-developed questionaire was used to collect data. 11 patients have additional data from in person interviews	Investigate sex and surgery satisfaction among MTF transsexual patients post SRS.	Twenty-seven of the 28 patients reported high satisfaction in their perceptions of the quality of their lives. Seventeen reported satisfaction in their employment after surgery. Twenty-one reported that SRS solved most of their emotional problems. However, some were, to a degree, **disappointed because of difficulties experienced postoperatively in adjusting satisfactorily as women both in their relationships with men and in living their lives generally as women**.		1. Retrospective study, no information before SRS, no before-after comparison. 2. Low response rate (28 out of 44 alive patients responded, 3 died), **selection/respondent bias.**	The first weakness is that it is retrospective and cross-sectional; no data obtained before SRS, no before-after comparison. The second weakness is the potential nonresponse bias, only 60% of the patients respondents. The non-respondents and the respondents might be systematically different, e.g. it is likely that patients who feel more satisfied were more likely to respond.

Johansson, A., Sundbom, E., Ho jerback, E., et. al. (2009).	Prospective	A prospective and longitudinal study of 42 patients: 5 years in the process and 2 years post surgery	To investigate outcome in terms of clinicians' and patients' evaluation of the process of sex reassignment.	The clinicians rated the global outcome as favorable in 62% of the cases, compared to 95% according to the patients themselves, with no differences between the subgroups. **The general evaluation of the surgery treatment** showed that of 33 patients (32 with genital surgery and one with only mastectomy), 22 (66.7%) were satisfied, seven (21.2%) were neither/nor, and four (12.1%) were dissatisfied	prospective and longitudinal		There is a large discrepancy between global outcome evaluation by clinicians and patients, with the patients reporting a much higher favorable rate (33%) than the clinicians. This shows the potential bias in self-reported satisfaction in patients, pointing to the necessarity of independent physician review. Also, 22 out of 33 (66.7%) patients were satisfied with the SRS treatment. So it is certainly not as consistent and overwhelmingly positive as the Assertion.
Hepp U, Klaghofer R, Burkhard-Kubler R, Buddeberg C. (2002).	Retrospective	In a retrospective study, 33 transsexual patients, 22 male-to-female transsexuals (MFTS), and 11 female-to-male transsexuals (FM-TS), were interviewed 53–121 months after their first referral to the psychiatric department of a university hospital. 25 patients had gone through surgical sex reassignment, while 29 were currently undergoing hormonal treatment.	physical and psychosocial well-being was satisfactory. Psychometric measures yielded normal values, with some pathological findings regarding personality traits. In the majority of patients, self- and observer-rating appraisals of gender-specific physical apearance were equally positive.			1. retrospective; 2. no control group; 3. small sample size.	The article is in German; I only reviewed the English abstract, which does not contain all the information on the outcome measures and statistical analysis, e.g. it is unclear whether the study measured outcome directly on " interpersonal relationships and improved social functioning." Similar as many studies in this literature, it has at least three methodological shortcomings: (1) retrospective; (2) no control group; (3) small sample size.
Ainsworth, T. & Spiegel, J. (2010)	duplicate						Reviewed in Assertion 1. Specific to this assertion, the paper has outcomes measured social functioning. But it suffers the same methodogical shortcomings (e.g. retrospective) as pointed notes on Assertion 1.
Smith et al. (2005)	duplicate						Reviewed in Supple. Assertion1. Specific to this assertion, this paper reported results on various measures of psychological functioning (e.g. shyness, anxiety, depression, negativism, sensitivity), but none of these are specific to "satisfaction with interpersonal relationships and improved social functioning"

Ettner Assertion # 6
continued from Ettner Assertion #5 …improvement in self-image and satisfaction with body and physical appearance… Ettner Report Para. 58.

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Lawrence, A. (2003).	Retrospective	A retrospective survey of 232 MTF patients using self-developed questionaire. (**The study design and patient sample is the same as in Lawrence 2005, 2006)**	To examined factors associated with satisfaction or regret following sex reassignment surgery	Participants reported overwhelmingly that they were happy with their SRS results and that SRS had greatly improved the quality of their lives. None reported outright regret and only a few expressed even occasional regret. Dissatisfaction was most strongly associated with unsatisfactory physical and functional results of surgery.The physical results of SRS may be more important than reoperative factors in predicting postoperative satisfaction or regret.		(1) retrospective; (2) low response rate of survey (32%)	The study has four post-SRS outcomes: (1) overall happiness with SRS result; (2) Improved QOL; (3) Regret; (4) Reversion (to living as a man after SRS). **None of these outcomes is specific to "improvement in self-image and satisfaction with body and physical appearance " as stated in the Asssertion.**
Weyers, S. et al. (2009)	Retrospective	A retrospective review of fifty transsexual women who had undergone SRS 6 months earlier were recruited in the Netherlands. Self-reported physical and mental health using the Short-Form-36 (SF-36) Health Survey; sexual functioning using the Female Sexual Function Index (FSFI).	To gather information on physical, mental, and sexual well-being, health-promoting behavior and satisfaction with gender-related body features of transsexual women who had undergone SRS.	Transsexual women function well on a physical, emotional, psychological and social level. With respect to sexuality, **they suffer from specific difficulties**, especially concerning arousal, lubrication, and pain.		(1) retrospective; (2) did not report the follow-up duration other than it is six months. (3) potential recruitment bias (50 out of 70 responded).	The results are not clear cut positive across all outcomes as stated in the Assertion. In addition, the paper has a number of shortcomings: (1) retrospective and cross-sectional, no before-after comparison; therefore no data supporting the Assertion of "**improvement** in self-image and satisfaction with body and physical appearance". (2) did not report the follow-up duration other than all patients had SRS at least six months ago. So it is the definition of long-term is unclear; (3) there is a potential recruitment bias (50 out of 70 responded. The paper stated “A study of the medical records of nonresponders and nonparticipants revealed that a volunteering bias may have been unlikely: we found non-significant differences in age, interval since surgery, psychiatric morbidity, operative techniques and complications between the study population and nonparticipating individuals. " However, such a justification is weak and doesn't consider the high possibility of unmeasured confounding.

Smith et al. (2005)	duplicate						Reviewed in Supple. Assertion 1. Specific to this assertion, this paper **reported evidence supporting** the assertion of "improvement in self-image and satisfaction with body and physical appearance"

Ettner Assertion # 7

continued from Ettner Assertion #5 ...and greater acceptance and integration into the family… Ettner Report Para. 58.

Name		Study-Design/Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Lobato, M., Koff, W, Manenti, C., Seger, D. et al. (2006).	Retrospective	Follow-up study of 19 patients post SRS in Brazil; outcomes are obtained via questionnaire. Follow-up time is between 1-2.5 years	To examine the impact of sex reassignment surgery on the satisfaction with sexual experience, partnerships, and relationship with family members in a cohort of Brazilian transsexual patients.	Sexual experience was considered to have improved by 83.3% of the patients, and became more frequent for 64.7% of the patients. For 83.3% of the patients, sex was considered to be pleasurable with the neovagina/neopenis. In addition, 64.7% reported that initiating and maintaining a relationship had become easier. The number of patients with a partner increased from 52.6% to 73.7%. **Family relationships improved in 26.3% of the cases, whereas 73.7% of the patients did not report a difference.**		1. retrospective, 2. small sample size; 3: respondent rate is 73% (19 out of 26) - potential respondent bias	The paper found that "Family relationships improved in 26.3% of the cases, whereas 73.7% of the patients did not report a difference. " Therefore, **it is not a clear cut as stated in the Assertion**. The paper in fact showed that that **family relationship only improved in 1/4 of the patients.**

Ettner Assertion # 8

Studies have also shown that gender-affirming surgery **improves patients' abilities to initiate and maintain intimate relationships**. Ettner Report Para. 59.

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Lawrence (2005), Sexuality after male-to-female sex reassignment surgery. Archives of Sexual Behaviour, 34(2), 147–166. https://doi.org/10.1007/s10508-005-1793-y. (full cite not provided in appendix, found in WPATH reference list)	Retrospective	A retrospective survey of 232 MTF patients using self-developed questionaire	To investigate sexual behaviors and attitudes before and after MTF SRS	Description of the prevalence of various sexual behaviors and attitudes before and after MTF SRS.		1. retrospective; 2. low response rate of survey (32%)	This is a descriptive study, not a comparative study. It describes the prevalence of various sexual behaviors and attitudes before and after MTF SRS. But there is no formal before-after comparative analysis or conclusions. There is no result in this paper supporting the Assertion "Studies have also shown that gender-affirming surgery improves patients' abilities to initiate and maintain intimate relationships." Two weakness of the methodology: (1) retrospective; (2) low response rate of survey

Imbimbo, C., Verze, P., Palmieri, A., et al (2009).	Restrospective	163 male patients who had undergone gender-transforming surgery at the authors' institution were requested to complete a patient satisfaction questionnaire, 139 (85%) responded. Follow-up time is 12-18 months.	To arrive at a clinical and psychosocial profile of male-to-female transsexuals in Italy through analysis of their personal and clinical experience and evaluation of their **postsurgical satisfaction levels SRS**.	Almost all of the patients were satisfied with their new sexual status and expressed no regrets. **Statement related to the Assertion** "75% had a more satisfactory sex life after SRS"	1. restrospective survey, 2. the questionnaire was not validated due to the inexistence of a validated questionnaire in this area of study.	This is a descriptive study, not a comparative study. There is no before-after SRS comparison. Because it is a retrospective survey, it is subject to recall bias. There is **no direct evidence in the paper** that supports the Assertion of "gender-affirming surgery improves patients' abilities to initiate and maintain intimate relationships." The only relevant statement of "75% had a more satisfactory sex life after SRS" is vague and can be purely on physical functions.
Klein, C. & Gorzalka, B. (2009)	Literature review	Outline cross-sex hormone therapy and SRS techniques, discuss the potential roles of cross-sex hormone therapy and SRS on sexual function, and **review the peer-reviewed literature** published in English on postoperative sexual functioning in MtF and FtM transsexuals.	To discuss the potential impact of cross-sex **hormone therapy and SRS** on sexual function and to summarize the published empirical research on postsurgical sexual functioning in male-to-female (MtF) and female-to-male (FtM) transsexuals.	Transsexuals appear to have adequate sexual functioning and/or high rates of sexual satisfaction following SRS. Further research is required to understand fully the effects of varying types and dosages of cross-sex hormone therapies and particular SRS techniques on sexual functioning.		This is a qualitative literature review on sexual function post both hormone therapy and SRS. The authors noted "From this review it is clear that there is **great variability in the sexual functioning of post-operative transsexuals**, and that **no clear outcome can be predicted with respect to whether surgery will have a positive or negative impact on sexual function**." Therefore, it does not support the Assertion.
De Cuypere, G., T'Sjoen, G., Beerten, R., Selvaggi, G., et al. (2005).	Restrospective	A long-term follow-up study of 55 transsexual patients (32 male-to-female and 23 female-to-male) post-sex reassignment surgery (SRS) in the Nethelands	To evaluate sexual and general health outcome post SRS	After SRS, the transsexual person's expectations were met at an emotional and social level, but **less so at the physical and sexual level** even though a large number of transsexuals (80%) reported improvement of their sexuality.	1. retrospective; 2. low response rate (55 out of 107, 51.4%)	The authors concluded that "After SRS, the transsexual person's expectations were met at an emotional and social level, but less so at the physical and sexual level even though a large number of transsexuals (80%) reported improvement of their sexuality." So this at least shows the effect of SRS on relationships is not a clear cut as stated by the relevant Assertion
Lawrence (2006) (full cite not provided in appendix, found on Cornell website)	Retrospective	A retrospective survey of 232 MTF patients using self-developed questionaire. (The study design and patient sample is the same as in Lawrence 2005, 2006)	To examine preoperative preparations, complications, and physical and functional outcomes of male-to-female sex reassignment surgery (SRS)	Satisfaction with most physical and functional outcomes of SRS was high; participants were least satisfied with vaginal lubrication, vaginal touch sensation, and vaginal erotic sensation.	(1) retrospective; (2) low response rate of survey (32%)	The paper focuses on surgical complications and physical and functional outcomes. It doesn't provide direct data supporting this Assertion (Suppl. 8) on "patients' abilities to initiate and maintain intimate relationships. "
Lobato et al., 2006	duplicate					Reviewed in Supple. Assertion 7. Relevant to this Assertion (Suppl. 8), the paper found "64.7% reported that initiating and maintaining a relationship had become easier" Therefore, **it is not a clear cut as stated in the Assertion**

Jarolím et al. (2009)	duplicate	Reviewed in Supple. Assertion 1. It focuses on describing the surgical details and outcomes of sexual functions and complications, **none of which is specific to this Assertion** (Suppl. 8) on "patients' abilities to initiate and maintain intimate relationships. "
Smith et al. (2005)	duplicate	Reviewed in Supple. Assertion 1. This paper reported results on various measures of psychological functioning (e.g. shyness, anxiety, depression, negativism, sensitivity), but none of these are specific to the Assertion on "patients' abilities to initiate and maintain intimate relationships. "
Rehman et al. (1999)	duplicate	Reviewed in Supple Assertion 5. The paper **did not report any outcome specific to this Assertion**, namely " patients' abilities to initiate and maintain intimate relationships. "

Ettner Assertion # 9

Multiple long-term studies have confirmed these results. Ettner Report Para. 60.

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Johansson et al. (2010) doi: 10.1007/s10508-009-9551-1 (not listed in the appendix – found from the Cornell website)	Retrospective	Follow-up study of 42 Swedish transexuals (5 year in the process or 2-year post SRS). Outcomes are extracted via semi-structured interview.	To evaluate the outcome of sex reassignment as viewed by both clinicians and patients, with an additional focus on the outcome based on sex and subgroups.	The clinicians rated the global outcome as favorable in 62% of the cases, compared to 95% according to the patients themselves, with no differences between the subgroups. Based on the follow-up interview, more than 90% were stable or improved as regards work situation, partner relations, and sex life, but 5–15% were dissatisfied with the hormonal treatment, results of surgery, total sex reassignment procedure, or their present general health.		1. Retrospective 2. Attrition (42 out of 60 patients, 70% respondent rate)	Besides the common methodological shortcomings (retrospective and attrition), the paper found a large discrepancy between clinician and patients assessment (patients more positive than clinician), similar to the previous finding by the same author (Johansson et al. 2009). There are 5–15% patients who were dissatisfied with the hormonal treatment, results of surgery, total sex reassignment procedure, or their present general health. So the results are not as clear cut as stated in the Assertion.

Vujovic S, Popovic S, Sbutega-Milosevic G, Djordjevic M, Gooren L. (2009).	Retrospective	A retrospective review of subjects applying for sex reassignment in 20 years in Serbia. 147 patients in total	To describe a transsexual population seeking sex ressignment treatment in Serbia	The relatively young age of those applying for sex reassignment and the sex ratio of 1:1 distinguish the population in Serbia from others reported in the literature.		Retrospective	This paper focuses on describing the demographic characteristics of transexual population. **It does not specify any outcome measures (e.g. QoL, satistification)**. The only related statements about post-SRS outcomes are (1) "In our population, there were no cases who regretted sex reassignment treatment", and (2) "After sex reassignment surgery, they feel more attractive and confident to live the lives of women", but **there is no data supporting the claim.**
Weyers et al. (2009)		duplicate					Reviewed in Supple Assertion 6. Related to this specific Assertion, a few shortcomings: (1) did not report the follow-up duration other than all patients had SRS at least six months ago. So it is not clear how long term is the follow up. (2) The results are not clear cut positive across all outcomes as stated in the Assertion, e.g. the paper found " **they suffer from specific difficulties, especially concerning arousal, lubrication, and pain**"
Hepp et al. (2002)		duplicate					Reviewed in Supple. Assertion 5. It is in German and I only reviewed the English abstract. It is indeed a long-term follow-up (53–121 months). Besides the methodological shortcomings, it is unclear whether the study measured outcome directly on all the outcomes stated in Supple Assertions 5-8
Imbimbo et al. (2009)		duplicate					Reviewed in Supple. Assertion 8. Besides the shortcomings that were pointed out, the follow-up time is 12-18 months, which is unclear to fit into the definition of "long term"
Lobato et al. (2006)		duplicate					Reviewed in Supple. Assertion 5, follow up time is 1 to 2.5 years, whether that fits in the definition of long-term is unclear. As reviewed before, the results are not clearcut.

Ettner Assertion # 10

Research shows that the risk of suicide can be significantly diminished with prompt and effective treatment. Ettner Report Para. 77.

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes

Bauer, G., Scheim, A., Pyne, J. et al. (2015).	Respondent-driven sampling survey	Respondent-driven sampling (RDS) survey of 380 transgender people on suicide outcomes. Among these, 110 subjects are used in the counterfactual analysis related to medical transition.	To identify intervenable factors (related to social inclusion, transphobia, or sex/gender transition) associated with reduced risk of past-year suicide ideation or attempt, and to quantify the potential population health impact.	Large effect sizes were observed for this controlled analysis of intervenable factors, suggesting that interventions to increase social inclusion and access to medical transition, and to reduce transphobia, have the potential to contribute to substantial reductions in the extremely high prevalences of suicide ideation and attempts within trans populations. Such interventions at the population level may require policy change.	Statistically rigorous (one of the most statistically advanced studies in the field); used the concept of counterfactual modeling (standard causal inference terminology),	no control of the different characteristics between intervention and intervention group.	Relevant to the assertion, the paper states (page 12 of 15) "The process of medically transitioning overall was more complex …We did not observe an increased risk in this sub-group among those who completed a medical transition (RR = 0.51; 0.07, 3.74)". Exact number is reported on the second page of Table 4 (page 11 of 15), top row: among the subjects who completed medical transition (100 Subjects), the relative risk of suicidal attempt is 0.51 with 95% confidence interval (0.07, 3.74), which is not statistically significant. So **the claim of "significantly diminished" is not supported by the statistical analysis**.

Ettner Assertion # 11

Moreover, gender dysphoric individuals have a profound discomfort or disgust of their genitalia. Without effective treatment as outlined above, this often leads to attempts at surgical self-treatment (SST), which can result in lasting physical trauma or death. Ettner Report Para. 78.

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes
Brown, G. & McDuffie, E. (2009)	Survey	Information of respondent of a survey to transgender inmates in prison systems in the U.S. was qualitatively analyzed and summarized.	Provide a summary of health care policies on transgender inmates in prison systems in the United States.	Note: not relevant to the assertion		Low response rate of the survey (46 respondents, 6 states no response).	The assertion that "this **often** leads to attempts at surgical self-treatment (SST)" is **not supported by the cited paper**, which states "**this rarely occurs in the community absent psychosis** " and reports (but without detailed data) that "the authors have firsthand knowledge of completed autocastration and/or autopenectomy in six facilities in four states." The paper contains no information on the **frequency** of this behavior. During my review, I found Brown (2014) reported that "Five percent of inmates reported that they had attempted (2%) or completed (3%) autocastration while incarcerated." **This percentage supports the occurance is rare rather than often.**

Ettner Assertion # 12

In addition, a systematic meta-analysis on publications performed by German researchers included 1,100 post-surgery participants. Seven different measures of quality of life were employed. The researchers concluded that gender-affirming surgery positively affects well being, sexuality, and quality of life in general. Ettner Report Para. 116.

Name	Study-Design	Methodology	Aim/Objective	Conclusions	Strengths	Limitations	Notes

Weinforth, et al., 2019.	Literature review	Review of 13 studies (2 prospective, 11 restrospective)	Provide a review of the currently available data on QoL after male-to-female SRS		Mixed prospective and retrospective designs, different questionaries used for outcome. Most are short term outcomes. No meaningful summaries.	This is a literature review, not meta-analysis, no formal statistical synthesis, only qualitative summaries. Mixed prospective and retrospective designs, different questionaries used for outcome, most studies are on short term outcomes. No meaningful summaries. Only two studies are prospective (Cardoso da Silva et al.; Lindqvist et al.), which have been reviewed in Assertion 1 and both showed mixed to negative results. The other 9 studies are all of low quality (retrospective/cross sectional studies, small sample sizes, no causal inference methods – no formal confounding adjustment other than regression. The authors acknowledges "prospective studies with standardized methods of assessing quality of life and with longer follow-up times would be desirable.